

DTIC FILE COPY

AGARD-AG-314

AGARD-AG-314

AD-A225 966

AGARD

ADVISORY GROUP FOR AEROSPACE RESEARCH & DEVELOPMENT

7 RUE ANCELLE 92200 NEUILLY SUR SEINE FRANCE

AGARDOGRAPH No. 314

Analysis, Design and Synthesis Methods for Guidance and Control Systems

(Les Méthodes d'Analyse de Conception et de Synthèse
pour les Systèmes de Guidage et de Pilotage)

DISTRIBUTION STATEMENT A

Approved for public release
Distribution Unlimited

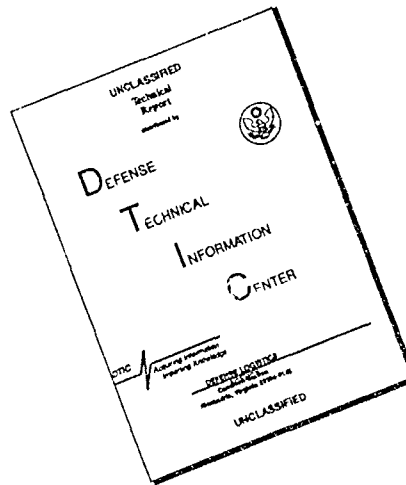
NORTH ATLANTIC TREATY ORGANIZATION



REPRODUCED BY
U.S. DEPARTMENT OF COMMERCE
NATIONAL TECHNICAL
INFORMATION SERVICE
SPRINGFIELD, VA 22161

*Original contains color
plates: All DTIC reproductions
will be in black and
white*

DISCLAIMER NOTICE



THIS DOCUMENT IS BEST
QUALITY AVAILABLE. THE COPY
FURNISHED TO DTIC CONTAINED
A SIGNIFICANT NUMBER OF
PAGES WHICH DO NOT
REPRODUCE LEGIBLY.

AGARD-AG-314

NORTH ATLANTIC TREATY ORGANIZATION
ADVISORY GROUP FOR AEROSPACE RESEARCH AND DEVELOPMENT
(ORGANISATION DU TRAITE DE L'ATLANTIQUE NORD)

AGARDograph No.314

Analysis, Design and Synthesis Methods for Guidance and Control Systems

(Les Méthodes d'Analyse de Conception et de Synthèse pour
les Systèmes de Guidage et de Pilotage)

Edited by

Professor C.T.Leondes
University of California
5532 Boelter Hall
Los Angeles, CA 90024
United States

Accession For	
NTIS CRA&I	<input checked="checked" type="checkbox"/>
DTIC TAB	<input type="checkbox"/>
Unannounced	<input type="checkbox"/>
Justification	
By	
Distribution /	
Availability Codes	
Dist	Avail and/or Special
A-1	



This AGARDograph has been prepared at the request of the
Guidance and Control Panel of AGARD.

The Mission of AGARD

According to its Charter, the mission of AGARD is to bring together the leading personalities of the NATO nations in the fields of science and technology relating to aerospace for the following purposes:

- Recommending effective ways for the member nations to use their research and development capabilities for the common benefit of the NATO community;
- Providing scientific and technical advice and assistance to the Military Committee in the field of aerospace research and development (with particular regard to its military application);
- Continuously stimulating advances in the aerospace sciences relevant to strengthening the common defence posture,
- Improving the co-operation among member nations in aerospace research and development;
- Exchange of scientific and technical information;
- Providing assistance to member nations for the purpose of increasing their scientific and technical potential;
- Rendering scientific and technical assistance, as requested, to other NATO bodies and to member nations in connection with research and development problems in the aerospace field.

The highest authority within AGARD is the National Delegates Board consisting of officially appointed senior representatives from each member nation. The mission of AGARD is carried out through the Panels which are composed of experts appointed by the National Delegates, the Consultant and Exchange Programme and the Aerospace Applications Studies Programme. The results of AGARD work are reported to the member nations and the NATO Authorities through the AGARD series of publications of which this is one.

Participation in AGARD activities is by invitation only and is normally limited to citizens of the NATO nations.

The content of this publication has been reproduced directly from material supplied by AGARD or the authors.

Published June 1990

Copyright © AGARD 1990
All Rights Reserved

ISBN 92-835-0566-2



*Printed by Specialised Printing Services Limited
40 Chigwell Lane, Loughton, Essex IG10 3TZ*

Preface

It was in the 18th century that Harrison received £10,000 for the invention of the chronometer which, in combination with the sextant, provided the answer to the challenge of the "quest for longitude" for the navigation of ships at sea. When one reflects on today's technology in guidance and control against the background of events such as this and others from earlier times, one can, perhaps, more fully appreciate the almost incredible things that are possible in guidance and control systems with today's high technology, capabilities which we sometimes almost take for granted.

In any event, the modern era of guidance and control systems can perhaps trace its origins to Professor C.S. Draper's pioneering work in the development of inertial guidance and control systems as embodied in the SPIRE system of the early 1950s. That system seemed to be about as big as a desk and weighed about a ton. The advent of the conception and development of the 2 degree of freedom gyro in the latter part of the 1950s was an important stepping stone to inertial guidance systems of much smaller size, weight, and volume. This eventually made it possible for virtually every military, commercial, and larger general aviation aircraft to have an inertial guidance system.

The advent of integrated electronic circuits, airborne digital computers, Kalman filter techniques and other advances presaged rapid advances in capability and performance through the 1960s and 1970s. Now in the 1980s further powerful technology advances including RLGs (Ring Laser Gyros), fiber optic gyros, NAVSTAR/GPS (Global Positioning Satellite), VLSI (Very Large Scale Integrated) electronics, and others presage even more powerful advances in capability and performance, while at the same time make such systems increasingly cost effective.

This NATO AGARDograph captures the spirit of such momentous developments. It is structured in eight parts. These are

Part I — Integrated Guidance and Control Systems

Part II — NAVSTAR/GPS Systems

Part III — Optical Gyroscope Guidance and Control Systems

Part IV — Integrated Communication and Navigation Systems

Part V — Integrated Navigation/Flight Control Systems

Part VI — Civil Aircraft Navigation and Traffic Control

Part VII — Special Topics

Part VIII — Land Navigation Systems

It should prove to be an invaluable reference source throughout the NATO community for many years to come, thanks to the splendid contributions by the co-authors.

Thanks are due to many people in the development of this volume. First, the editor would like to thank Mr Ken Peebles, Panel Chairman, and members of the Guidance and Control Panel who expressed their desire for and supported such a volume. Next, the Guidance and Control Panel Executive, A. Rocher, Colonel, FAF, and James Ramage, the Panel US Coordinator, provided outstanding support. In addition, my secretary, Ruth Greene, showed tremendous dedication to the time consuming efforts she expressed so cheerfully to make this volume possible.

Finally, the editor would like to express his sincere appreciation to his many colleagues throughout the NATO community who, in one way or another, contributed to his continually evolving concepts about guidance and control systems.

C.T. Leondes

Préface

Au 18ème siècle un certain M.Harrison a reçu la somme de 10000 livres en récompense de l'invention du chronomètre, lequel, associé au sextant, a fourni la solution au problème de "la recherche de la longitude" pour la navigation maritime. En comparaison, et grâce aux techniques de pointe, les technologies employées dans le domaine du guidage et du pilotage aujourd'hui offrent des possibilités presque inimaginables, possibilités que l'on a trop souvent tendance à considérer comme banales.

Quoiqu'il en soit, il semble que l'ère moderne des systèmes de guidage et de pilotage trouve ses origines dans les travaux originaux du Professeur C.S.Draper sur les systèmes de guidage et de pilotage inertiels qui ont débouché sur le système SPIRE au début des années 1950. Le système en question était aussi volumineux qu'un bureau et pesait au moins une tonne. L'arrivée du gyroscope à deux degrés de liberté, vers la fin des années 1950, a servi de tremplin pour la réalisation de systèmes de guidage inertiels plus légers et plus compacts. Grâce à ces initiatives, pratiquement tous les aéronefs militaires, commerciaux et d'aviation générale modernes sont équipés de centrales inertielles.

L'arrivée de circuits électroniques intégrés, de calculateurs numériques embarqués, de techniques de filtrage Kalman et d'autres innovations technologiques ont permis de rapides progrès en capacité et en performances tout au long des années 1960 et 1970. Les progrès technologiques importants qui ont suivi au cours des années 1980 dans les RLG (le gyrolaser en anneau) les gyroscopes à fibre optique, le NAVSTAR/GPS, (système global de localisation, les circuits VLSI (intégration de l'électronique à très grande échelle) et d'autres encore laissent prévoir des innovations plus performantes qui permettront d'augmenter les capacités des systèmes tout en réduisant le coût.

Cette AGARDographie s'inspire de l'esprit de telles innovations et est organisée en huit sections.

Section I — Les systèmes intégrés de guidage et de pilotage

Section II — Les systèmes NAVSTAR/GPS

Section III — Les systèmes de guidage et de pilotage à gyroscope optique

Section IV — Les systèmes intégrés de télécommunications et de navigation

Section V — Les systèmes intégrés de navigation/commandes de vol

Section VI — La navigation et le contrôle de la circulation des avions civils

Section VII — Questions particulières

Section VIII — Systèmes de navigation terrestre

Grâce aux contributions des co-auteurs, cet ouvrage devrait se révéler très utile comme source de référence pour la communauté de l'OTAN pendant les années à venir.

La liste des personnes qui ont contribué à la rédaction de ce volume est longue. Premièrement, l'auteur tient à remercier M.Ken Peebles, Président, et les membres du Panel AGARD du Guidage et du Pilotage qui ont souhaité la parution de ce volume et qui ont tout fait pour la faciliter. Ses remerciements vont ensuite à l'Administrateur du Panel du guidage et du pilotage, le Col. A.Rocher FAF, et M.James Ramage le coordonnateur du Panel pour les Etats-Unis pour leur soutien exceptionnel; sans oublier sa secrétaire, Ruth Greene, qui a fait preuve d'un dévouement et d'une bonne humeur extraordinaires dans le travail de secrétariat qui s'imposait.

Enfin, l'auteur tient à exprimer ses sincères remerciements à ses nombreux collègues au sein de la communauté de l'OTAN, qui ont contribué de bien des façons, à l'évolution de sa pensée sur les systèmes de guidage et de pilotage.

C.T.Leondes

Contents

	Page
Preface	iii
Préface	iv
	Reference
PART I — INTEGRATED GUIDANCE AND CONTROL SYSTEMS	
GPS Integrity Requirements for Use by Civil Aviation by A.K.Brown	I1
Integration of GPS and Strapdown Inertial Subsystems into a Single Unit by D.Buechler and M.Foss	I2
Application of Multifunction Inertial Reference Systems to Fighter Aircraft by C.A.Bedoya and J.M.Perdzock	I3
Kalman Filter Formulations for Transfer Alignment of Strapdown Inertial Units by A.M.Schneider	I4
Combining Loran and GPS — the Best of Both Worlds by P.Braisted, R.Eschenbach and A.Tiwari	I5
The Integration of Multiple Avionic Sensors and Technologies for Future Military Helicopters by A.J.Shapiro	I6
PART II — NAVSTAR/GPS (GLOBAL POSITIONING SATELLITE) SYSTEMS	
Techniques for Autonomous GPS Integrity Monitoring by B.W.Parkinson and P.Axelrad	II1
Modular Digital GPS Receivers by J.S.Graham, P.C.Ould and R.J.Van Wechel	II2
Integration of GPS/INS with Partitioned Filters by J.W.Diesel	II3
2-D and 3-D Characterizations of GPS Navigation Service by P.Massat, W.Rhodus and K.Rudnick	II4
Applications of Differential GPS by K.Hervig and H.Fjæreide	II5
An Analysis of GPS as the Sole Means Navigation System in US Navy Aircraft by G.Löwenstein, J.Phanos and E.C.Rish	II6
The Determination of PDOP (Position Dilution of Precision) in GPS by A.H.Phillips	II7
PART III — OPTICAL GYROSCOPE GUIDANCE AND CONTROL SYSTEMS	
Ring Laser Gyro Principles and Techniques by G.J.Martin	III1

	Page
Inertial Grade Fiber Gyros by G.A.Pavlath	III2
Use of a Three-Axis Monolithic Ring Laser Gyro and Digital Signal Processor in an Inertial Sensor Element by D.J.Weber	III3
Ring Laser Gyro Marine Inertial Navigation Systems by C.C.Remuzzi	III4

PART IV -- INTEGRATED COMMUNICATION AND NAVIGATION SYSTEMS

Distributed Control Architecture for CNI Preprocessors by V.R.Subramanyam and L.R.Stine	IV1
JTIDS Relative Navigation -- Principles, Architecture and Inertial Mixing by W.R.Fried	IV2
PLRS -- A New Spread Spectrum Position Location Reporting System by J.A.Kivett and U.S.Okawa	IV3
Enhancing PLRS with User-to-User Data Capability by J.A.Kivett and R.E.Cook	IV4
Observability of Relative Navigation Using Range-Only Measurements by A.M.Schneider	IV5
Integrated Strapdown Avionics for Precision Guided Vehicles by J.Richman, D.Haessig, Jr. and B.Friedland	IV6

PART V -- INTEGRATED NAVIGATION/FLIGHT CONTROL SYSTEMS

Integrated Navigation/Flight Control for Future High Performance Aircraft by R.E.Ebner and A.D.Klein	V1
Survivable Penetration by C.A.Bedoya, G.N.Maroon, W.J.Murphy and C.W.Chapoton, Jr.	V2

PART VI -- CIVIL AIRCRAFT NAVIGATION AND TRAFFIC CONTROL

Independent Ground Monitor Coverage of Global Positioning System (GPS) Satellites for Use by Civil Aviation by K.J.Viets	VI1
Analysis of the Integrity of the Microwave Landing System (MLS) Data Functions by M.B.El-Arini and M.J.Zeltser	VI2

PART VII -- SPECIAL TOPICS

Fault Detection and Isolation (FDI) Techniques for Guidance and Control Systems by M.A.Sturza	VII1
Control and Estimation for Aerospace Applications with System Time Delays by E.J.Knobbe	VII2
Overview of Omega Signal Coverage by R.R.Gupta and P.B.Morris	VII3
Omega Navigation Signal Characteristics by P.B.Morris and R.R.Gupta	VII4

	Page
Pointing Control System for the Teal Ruby Experiment by R.Rogers	VII15
The Potential for Digital Databases in Flight Planning and Flight Aiding for Combat Aircraft by J.Stone	VII16
The Assessment and Selection of Inertial Systems for Artillery by Y.K.Ameen and G.B.Symonds	VII17
Guidance and Control Techniques for Airborne Transfer Alignment and SAR Motion Compensation by J.L.Farrell	VII18
State Estimation for Systems Moving through Random Fields by D.E.Catlin and R.L.Geddes	VII19

PART VIII – LAND NAVIGATION SYSTEMS

A Low Cost Inertial/GPS Integrated Approach to Land Navigation by D.G.Harris	VIII1
A Land Navigation Demonstration Vehicle with a Color Map Display for Tactical Use by E.J.Nava, E.E.Creel, J.R.Fellerhoff and S.D.Martinez	VIII2
Satellite Navigation Systems for Land Vehicles by R.A.Dork and O.T.McCarter	VIII3
An Integrated System for Land Navigation by J.C.McMillan	VIII4

PART I

Integrated Guidance and Control Systems

GPS INTEGRITY REQUIREMENTS FOR USE BY CIVIL AVIATION

by

Alison K. Brown
 Navstar Systems Development
 18560 Lower Lake Rd
 Monument, CO 80132
 United States

SUMMARY

At the request of the Federal Aviation Administration (FAA), the Radio Technical Commission for Aeronautics (RTCA) established Special Committee 159 on September 20, 1985. The purpose of SC-159 was to prepare a "Minimum Aviation System Performance Standard" (MASPS) for the operation and use of the evolving Global Positioning System (GPS) in civil air navigation. To assist in preparing the MASPS, SC-159 formed an Integrity Working Group on April 22, 1986, chaired by the author, to investigate and report on civil integrity problems relating to GPS. The purpose of the Working Group was to establish GPS integrity monitoring requirements and to discuss suitable integrity monitoring techniques for civil aviation. The final report, was completed by the Working Group on June 3, 1987. This paper summarizes the Integrity Working Group recommendations to SC-159.

1. INTRODUCTION

Integrity was defined by the Working Group as the ability of a system to provide timely warnings to users when the system should not be used for navigation. To assure the safety of aircraft, a timely warning is required any time the performance of the navigation system fails to meet the accuracy requirements applicable to the particular phase of flight of the aircraft. Consequently, integrity warning time and accuracy threshold requirements vary with the phase of flight. Phases of flight considered by the Working Group were oceanic en-route, domestic en-route, terminal area and non-precision approach.

The Global Positioning System (GPS) was developed by the United States Department of Defense (DoD) to enhance the effectiveness of military missions and to reduce the proliferation of DoD radionavigation systems, thereby reducing costs. The extensive built-in self-checking and warning features of the GPS are adequate to meet military integrity requirements and to allow safe operation of DoD aircraft. However, more stringent safety requirements must be met for GPS to receive FAA approval for use by civil aviation in the National Airspace System.

A navigation system can receive FAA approval as either a sole-means navigation system or a supplemental navigation system for any phases of flight depending on its capabilities. A sole-means navigation system is one that may be used on an aircraft without any other means of navigation available. Conversely, a supplemental navigation system may only be used when a sole-means system is available should the supplemental system be unable to continue navigation. Generally, the sole-means navigation system does not provide the same level of accuracy as the supplemental system, but is continuously available as a safe, alternative means of navigation. While integrity is obviously a requirement for sole-means approval, integrity warnings are also required for supplemental approval to assure the safety of the aircraft when the supplemental system is providing primary navigation.

For sole-means navigation, the FAA is not only concerned with the integrity of the GPS navigation solution, but also in the continuity of the GPS service. To satisfy this requirement, sufficient redundant signal coverage must be provided so that the failure of any single element of the system (e.g. any GPS satellite) does not cause an interruption of service. With the planned GPS constellation of 18 satellites and three spares, this requirement is not met since redundant satellite coverage is not continuously available.

For supplemental navigation, there is no requirement on the continuity of the GPS service. However, the navigation system must provide integrity warnings when out of tolerance conditions exist. Although the GPS satellites broadcast health and status information which indicate the system integrity, this data is not updated promptly enough to meet FAA requirements for timely warnings.

For GPS to be used for supplemental or sole-means navigation in the civil airspace, the Working Group concluded that conventional GPS navigation must be augmented to provide integrity and, in the case of sole-means navigation, increased redundancy and coverage. A wide variety of integrated systems, processing techniques and improvements to the GPS were discussed by the Integrity Working Group to achieve these ends.

1.1 GPS NAVIGATION

The Global Positioning System (GPS) is a satellite based radionavigation system developed by the United States Department of Defense. The system operates through passive triangulation to four satellites, providing highly accurate three-dimensional position, velocity and time. Two classes of service are available, the Standard Positioning Service (SPS) for civil applications, and the Precise Positioning Service (PPS) for military and other authorized users. The PPS allows three-dimensional position to be determined to 16 m SEP (Spherical Error Probable). In the interests of national security, the SPS accuracy is degraded to 100 m 2dRMS, through the introduction of artificial Selective Availability (SA) errors on the GPS satellite signals.

The GPS is partitioned into three segments: the space segment which consists of the GPS satellites themselves; the control segment which tracks and maintains the satellites; and the user segment which includes the receiver used to navigate from the satellite signals.

The space segment is currently planned to include a constellation of 18 GPS satellites with three active (broadcasting) spares. The satellites will be launched into 12-hr orbits inclined at 55° , as shown in Figure 1, at an altitude of 20,183 km (10,898 nm). The orbits are chosen so that four satellites are continuously in view anywhere in the world. The full constellation of satellites is planned to be available in the mid 1990s.

The control segment consists of five monitor stations, a Master Control Station and three ground antennae. Each monitor station passively tracks the GPS satellites, accumulates satellite and meteorological data and transmits this information to the Master Control Station. The Master Control Station generates ephemeris and clock bias predictions and formulates the NAV message to be broadcast by the satellites. These messages are uploaded to the satellites, generally twice daily, using the ground antennae.

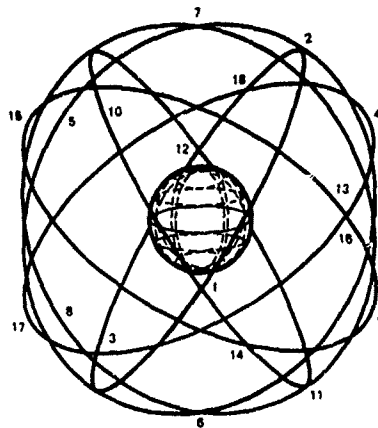


Figure 1. The 6-Plane 18-Satellite GPS Constellation

The user segment consists of the GPS navigation sets used to receive the GPS satellite signals. To navigate with GPS, a minimum of four GPS satellites must be tracked by the receiver. From the four Pseudo-Ranges (range + receiver clock offset) to the GPS satellites, the receiver can determine 3-dimensional position and GPS time. The Delta-Ranges (carrier doppler shifts) allow 3-dimensional velocity to be also computed.

1.2 GPS ERROR CHARACTERISTICS

The GPS navigation accuracy is a function of both the geometry of the four satellites being tracked and the precision of the Pseudo-Range measurements made from the satellite signals. The Position Dilution of Precision (PDOP) is the scaling effect of the satellite geometry between the satellite Pseudo-Range measurement and the GPS 3-dimensional position solution. The horizontal navigation accuracy provided by GPS can be computed by scaling the Pseudo-Range measurement errors by the Horizontal Dilution of Precision (HDOP). The horizontal 2dRMS accuracy of the GPS navigation solution is computed using equation (1-1).

$$\text{Horizontal Navigation Accuracy} = 2 \times \text{HDOP} \times \epsilon_{pr} \quad (1-1)$$

The SPS Pseudo-Range measurement accuracy (ϵ_{pr}) is derived from a combination of different error sources. These can be grouped into four categories: satellite clock and ephemeris errors; propagation uncertainties due to the atmosphere and reflected signals (multipath); receiver errors; and the artificial Selective Availability (SA) errors introduced for purposes of national security.

The satellite clock and ephemeris errors for the Block II GPS satellites are expected to be small, typically around 5 m, and so do not contribute significantly to the SPS error. Propagation uncertainties are primarily a function of the accuracy of the atmospheric compensation models. The residual error after compensation can be as large as 20 to 30 m in extreme cases. However, errors on the order of 5 m at night and 10 to 15 m during the day are more typical. The measurement error introduced by a SPS receiver is on the order of 15 m. In most cases, this can be significantly reduced through filtering.

The largest contributing error source for the SPS user is the Selective Availability (SA) errors deliberately introduced on the GPS satellite signals. Very little information has been released about the exact nature of the SA errors other than that they will be random and used to limit the SPS accuracy to 100 m 2dRMS. This is interpreted by the DoD, as in the Federal Radionavigation Plan, as a 95% figure. The SPS navigation errors may therefore be expected to exceed 100 m an appreciable fraction of the time. Tracing backward from the 100 m 2dRMS figure, the magnitude of the Selective Availability error on the Pseudo-Range signal will be about 30 m (RMS). When this is root-sum-squared with the other error sources it completely dominates the SPS error budget resulting in a total error (σ_{pr}) of about 33 m.

The 18-satellite GPS constellation always provides a minimum of four satellites in view, with HDOP typically in the range 1.5 to 1.7. Scaling this HDOP by the Pseudo-Range errors (σ_{pr}), as shown in equation (1-1), results in a total navigation accuracy of 100 m 2dRMS.

However, the 18-satellite constellation does not allow continuous world-wide navigation to this accuracy. Whenever the PDOP exceeds a threshold value of six (6), the user is considered to be in an area of degraded performance where GPS is not capable of un-aided navigation. These areas of degraded performance occur at regularly scheduled intervals at latitudes around 35° and 65° in both hemispheres and generally last for around 20 minutes. Figure 2. shows where these regions occur with the 18-satellite GPS constellation.

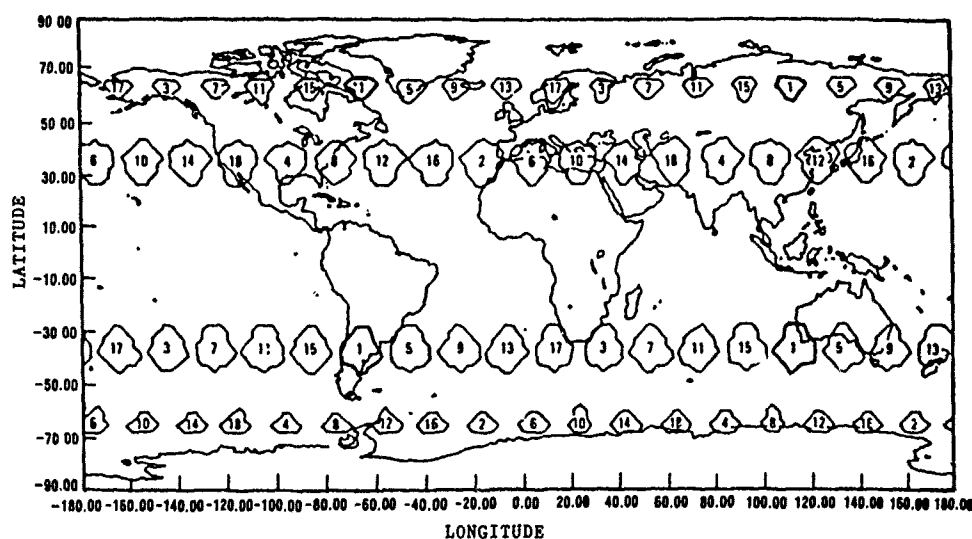


Figure 2. Areas of Degraded Performance with the 18-Satellite Constellation;
© 1981 IEEE

By selectively placing the three GPS satellite spares, it is possible to remove these areas of degraded performance over a limited region, for example the continental United States. However, there is only a 50% probability of having all 21 satellites operational. (There is a 98% probability of 18 satellites being operational.) As soon as a satellite failure occurs, the areas of degraded performance appear again over the United States. This coverage deficiency of the 18-satellite GPS constellation precludes the use of a stand-alone GPS receiver as a sole-means navigation system. Various methods of augmenting the GPS system were discussed by the Working Group to overcome this problem.

1.3 GPS SYSTEM INTEGRITY

The Global Positioning System has extensive built-in features and operating procedures to ensure the integrity of the navigation service. These include: equipment redundancy; communication error detection codes; estimation and prediction consistency checks; and operator qualification verification.

The Block II GPS satellites are designed to perform extensive self-checking with the provision to discontinue the ranging signal if internal failures are detected. In addition, the signals and data transmitted to users are continuously monitored by the Control Segment (with the exception of a small Pacific region west of Chile). However, a 15 to 20 minute delay exists between the anomaly occurrence and the earliest

indication of malfunction at the Control Segment. An additional hour is then required to deploy one of the ground antenna to the failed satellite and update the data transmitted to the user. The delay inherent in the Control Segment monitoring does not meet the FAA requirement of timely notification of system failures.

1.4 GPS INTEGRITY REQUIREMENTS

Under normal conditions, the GPS Standard Positioning Service will provide 100 m accuracy (2dRMS) to civil users. However, in the unlikely event of a GPS failure occurring, additional precautions must be taken by civil aviation GPS navigation sets to detect failures before the navigation errors exceed the allowable error threshold for a particular phase of flight.

To establish integrity alarm limits and time-to-alarm requirements for different phases of flight, the Working Group examined existing requirements already established for other navigation systems. Documents referenced included: the Department of Transportation Federal Aviation Administration Advisory Circular No. AC 90-45A; the Department of Defense and Department of Transportation Federal Radionavigation Plan (FRP)₂₀; and the Radio Technical Commission for Aeronautics Document No. RTCA/DO-180, "Minimum Operational Performance Standards for Airborne Area Navigation Equipment Using Multi-Sensor Inputs".

Requirements extracted from these documents are listed in Table 1-1. Where conflicting requirements occurred, the tighter or more stringent requirement was selected for GPS. Only horizontal navigation was considered because of problems with the compatibility of altitude reference systems. Radial alarm limits were selected as appropriate for GPS since the navigation errors are not dependent on the direction of the aircraft track.

Table 1-1 GPS Integrity Requirements

Phase of Flight	Oceanic En-Route	Domestic En-Route	Terminal Area	Non-Precision Approach
Alarm Limit	12.6 nmi	1.5 nmi	1.1 nmi	0.3 nmi
Time-to-Alarm	120 sec	60 sec	15 sec	10 sec

Because of the superior navigation accuracy normally possible with GPS, and with a view towards possibly reducing aircraft separation and obstacle clearance criteria in the future, the Working Group also established a set of goals for GPS integrity criteria based primarily on information from the Federal Radionavigation Plan₂₀. The goal criteria are listed in Table 1-2.

Table 1-2 GPS Integrity Goals

Phase of Flight	Oceanic En-Route	Domestic En-Route	Terminal Area	Non-Precision Approach
Alarm Limit	5 km	1 km	500 m	100 m
Time-to-Alarm	30 sec	30 sec	10 sec	6 sec

The different integrity monitoring techniques addressed by the Working Group, were studied to determine under which phases of flight GPS failures could be detected within the required time-to-alarm before they exceeded the accuracy alarm limit.

2. INTEGRITY MONITORING TECHNIQUES

The integrity monitoring techniques considered by the Working Group can be divided into two categories; internal methods and external methods. The different techniques studied are listed in Table 2-1. With internal methods, the GPS integrity can be determined using information provided by the aircraft sensors only. For example, redundant data inside the GPS receiver may be used, or aiding data supplied to the receiver from sensors such as a barometric altimeter or an inertial navigation system (INS). Using external methods the GPS signals are monitored in real-time through a network of ground monitoring stations. A variety of communication media were considered for disseminating the GPS integrity data to users.

Each of the integrity monitoring techniques addressed were analyzed to determine over which phases of flight they would be suitable using the integrity criteria defined in Tables 1-1 and 1-2. If appropriate, the monitoring techniques were also studied to determine whether they supplied sufficient redundancy for GPS to be used for sol-means navigation.

Table 2-1 GPS Integrity Monitoring Techniques

Internal Methods	External Methods
Receiver Autonomous Integrated Systems - GPS/Baro-altitude - GPS/INS/IRS/AHRS - GPS/LORAN-C - GPS/Omega - GPS/Multi-sensor FMCS - GPS/VOR-DME/RNAV	GPS Integrity Channel (GIC) - Ground-Based Communication - Satellite Communication Differential GPS

2.1 RECEIVER AUTONOMOUS INTEGRITY MONITORING

With Receiver Autonomous Integrity Monitoring (RAIM), the GPS receiver makes use of redundant information from the GPS satellites, or other sensors, as a check on the integrity of the navigation solution. A variety of different self-contained monitoring algorithms are possible, 7 8.

Figure 3. illustrates a simple 'snapshot' approach to GPS failure detection. In this case five GPS satellites are visible. Depending on which four GPS satellites are used for navigation, there are five possible different navigation solutions as shown in Figure 1., Case 1. All five solutions are scattered due to the normal GPS system errors. However, in the case shown, all five navigation solutions lie within 100 m of the true location of the aircraft. If the differences between the five navigation solutions are compared, none will exceed 200 m, the normal error spread to be expected using the GPS.

In Figure 3., Case 2, a failure is assumed to have occurred in satellite #5 causing all the navigation solutions using this satellite to be in error by differing amounts. If the differences between the navigation solutions are now compared, some will exceed the expected 200 m allowable level, indicating that a satellite failure has occurred. Since one of the navigation solutions does not contain the failed satellite, this method always ensures that one solution is correct and must be near the true location of the aircraft.

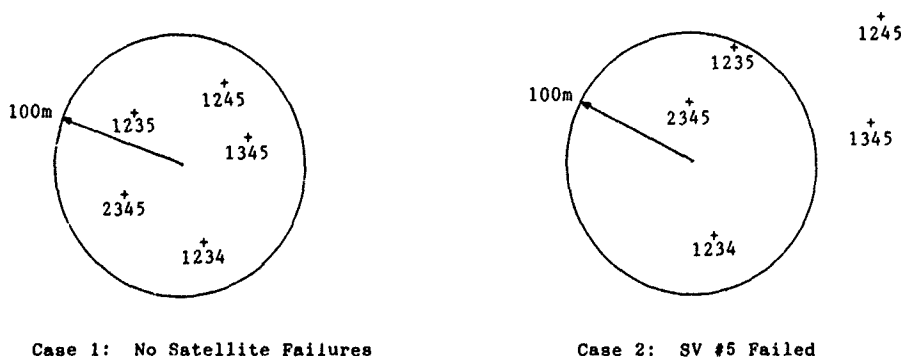


Figure 3. Possible Navigation Solutions with 5 GPS Satellites

For RAIM to be possible using this 'snapshot' approach, a navigation solution must always be possible, even when a satellite has failed. This reduces to a satellite geometry condition that all (N-1) subsets of satellites out of N visible satellites must give a PDOP sufficient to assure the navigation accuracy associated with the particular phase of flight of the aircraft.

The 'snapshot' approach can be improved by using Kalman Filter techniques, which take account of parameters such as doppler measurements and clock stability in the failure detection scheme. However, there is still insufficient redundancy provided by the 18 satellite GPS constellation to allow RAIM to be continuously effective. To provide continuous GPS integrity using this method, the satellite constellation must be augmented. Increasing the GPS constellation to 24 satellites would provide sufficient redundancy for RAIM to be effective. Augmenting the 18-satellite constellation with geostationary satellites would also provide sufficient redundancy. Two geostationary satellites located over the United States would provide integrity coverage over CONUS. Five geostationary satellites would be sufficient to provide world-wide coverage.

The level of integrity that can be provided through RAIM under good geometry conditions is primarily a function of the Selective Availability errors. Simulations have shown that the minimum alarm level that can be set in the presence of Selective Availability, without an excessive alarm rate, is around 300 m. From the integrity requirements listed in Tables 2-1 and 2-2, RAIM would be suitable to meet all the integrity requirements and goals, were sufficient satellite coverage available, except for the non-precision approach goal of 100 m.

2.2 INTEGRATED SYSTEMS

By integrating GPS with other navigation systems, it is possible to achieve highly accurate navigation performance while ensuring navigation integrity. In the interim period before GPS is fully operational, or to supplement the coverage provided by the 18 satellite constellation, integrated GPS navigation systems will prove effective for both supplemental and sole-means navigation. Integrating GPS with other navigation sensors provides additional redundant data which may be used for integrity monitoring. The additional data, in some cases, also increases the coverage provided by the GPS service and allows sole-mean navigation to be possible for the hybrid navigation system.

Some of the navigation systems addressed are already certified for sole-means navigation for some phases of flight. GPS may obviously be used as a supplemental navigation system with these, using the navigation data from the sole-means system as a check on the integrity of the GPS navigation solution. This will allow errors to be detected before they exceed the allowable error range for the particular phase of flight, should a failure occur. However, under normal operating conditions, the aircraft operator may take advantage of the superior navigation performance provided by GPS.

2.2.1 Baro-Altitude Aiding

The Barometric Altimeter has a long history as a cockpit instrument. In its modern form, the altitude can be provided digitally as an aiding sensor to the GPS receiver. Typically, the instrument errors of the barometric altimeter can be held within 200 ft. However, the barometric altimeter measures pressure altitude and thus is subject to meteorological vagaries in relating true altitude to pressure altitude. These differences can be quite large, and when not compensated for will dominate the intrinsic instrument errors. For example, it would not be unusual for pressure altitude to differ from true altitude by 1000 ft or more over long flights across the continent or ocean. Also, pressure altitude, even when corrected by a reporting station, can be in error by a few hundred feet at a different altitude and at a location a few tens of miles away.

The principle of baro-altitude aiding for integrity monitoring is similar to that described for RAIM where the altitude data is used as an additional GPS satellite Pseudo-Range measurement. This form of aiding is only applicable when less than five satellites are in view and RAIM alone cannot be effective. The baro-altitude aiding data is not as accurate as an additional Pseudo-Range measurement from a GPS satellite, so the same level of integrity possible with RAIM cannot be ensured. However, the use of baro-altitude aiding does provide additional redundancy and increases the navigation coverage supplied by GPS.

Further study is in progress, but it appears likely that baro-altitude aiding would allow the use of GPS as a sole-means navigation service for en-route oceanic navigation with the 18 satellite constellation. Because of the limitations of the baro-altitude accuracy, the integrity provided will probably not be sufficient for other phases of flight.

2.2.2 GPS/Inertial Integrated Systems

Inertial navigation systems (INS) are relative, not absolute, position sensors and so the navigation accuracy deteriorates with time. The inertial errors are generally characterized as a linear drift in position with a superimposed Schuler oscillation. In an integrated GPS/INS system the GPS data is used to calibrate the INS, while the INS is used to monitor the integrity of the GPS navigation. The INS cannot detect absolute position errors, but may be used to monitor against a slow drift occurring in the navigation solution. In conjunction with RAIM, when five satellites are available, this is an effective technique for monitoring the GPS integrity throughout periods with poor satellite coverage. The problem then reduces to comparing the inertial drift rate to the GPS error rate over the periods of time when RAIM is not effective with the 18 satellite constellation.

With the 18 satellite constellation, the maximum time that less than five satellites are in view is 63 minutes. In Table 2-2, the integrity requirements and goals shown in Tables 1-1 and 1-2 are equated to drift rates over a one hour period. To assure that the navigation error cannot exceed the alarm limit during the period that RAIM is not effective, the INS must detect error rates that exceed these limits.

Table 2-2 Error Rate Requirements for INS Integrity Monitoring

Phase of Flight	Oceanic En-Route	Domestic En-Route	Terminal Area	Non-Precision Approach
INTEGRITY REQUIREMENTS				
Drift Rate	6.4 m/s	0.8 m/s	0.6 m/s	0.15 m/s
Time-to-Detect	3720 sec	3660 sec	3615 sec	3610 sec
INTEGRITY GOALS				
Drift Rate	1.4 m/s	0.28 m/s	0.14 m/s	0.03 m/s
Time-to-Detect	3630 sec	3630 sec	3610 sec	3606 sec

From simulation results, a commercial 2 nm/hr INS integrated with a GPS receiver can meet the integrity requirements for en-route ocean and domestic phases of flight. The improved performance possible with an integrated GPS/INS system will meet the integrity and redundancy requirements for the sole-means en-route domestic phase of flight, and the goals established from future requirements for en-route oceanic navigation. Inertial navigation is already a certified sole-means of navigation for en-route oceanic phases of flight.

2.2.3 GPS/LORAN-C

Integrating a GPS and a LORAN receiver has the potential of providing a hybrid navigation system with superior performance than either system alone. Both GPS and LORAN-C suffer to differing extents from a lack of coverage. The 18 satellite GPS constellation provides world-wide coverage with the exception of localized areas which appear at different times of day. The LORAN-C network covers the majority of the United States with the exception of the mid-continent gap. Both systems use ranging techniques for navigation; GPS Pseudo-Ranges from satellites and LORAN-C Time-Delays (TD) from chains of ground transmitters. In principle, both the Pseudo-Ranges and TD measurements could be processed in an integrated solution using the redundant measurements for integrity monitoring similar to RAIM. Since the nominal accuracy of LORAN-C is 0.25 nmi, this hybrid navigation system will meet all the integrity requirements and goals listed in Tables 1-1 and 1-2, except for the non-precision approach goal of 100 m. The extended coverage provided by this hybrid should allow sole-means navigation to be possible for en-route domestic and terminal navigation phases of flight.

2.2.4 GPS/Omega

Like LORAN-C, Omega is also a ground-based radionavigation service. The Omega coverage extends over most of the world. However, the Omega navigation service only provides position to 2-4 nmi which significantly limits the use of the Omega ranges to monitor the GPS integrity or extend the GPS coverage. The only integrity requirements that can be met using Omega data for monitoring are the en-route oceanic requirements. Since Omega is already certified as a sole-means of navigation for this phase of flight, an integrated GPS/Omega receiver does not extend the range of operation of Omega. The GPS navigation set may be used as a supplemental navigation system for oceanic en-route phases of flight.

2.2.5 GPS/Multi-Sensor FMCS

The integration of GPS into the Flight Management Control System (FMCS) can be viewed simply as the introduction of another radionavigation sensor. As such, the GPS receiver must meet the requirements established in RTCA/DO-187, "Minimum Operational Performance Standards for Airborne Area Navigation Equipment using Multi-Sensor Inputs". The integrity requirements established in paragraph 2.2.1.11 of this document specify that the equipment should monitor itself for degraded performance and should annunciate degraded operation. This could be achieved by a variety of methods, for example RAIM. Following the guidance in DO-187 will allow supplemental use of GPS as a sensor input to an FMCS.

2.2.6 GPS/VOR/DME-RNAV

To make use of VOR/DME data to monitor the GPS navigation integrity, the raw range and bearing VOR/DME data must be converted into latitude and longitude coordinates. With the help of a navigation data-base, a VOR/DME Area Navigation Computer (RNAV) can convert these polar coordinates to latitude and longitude, as described in RTCA/DO-180. The output of this RNAV computer can be used as a truth reference to monitor the

integrity of the GPS navigation solution. This would allow supplemental use of GPS with an RNAV computer.

2.3 GPS INTEGRITY CHANNEL (GIC)

With a GPS Integrity Channel (GIC), a ground-based GPS monitoring system is used to track the GPS signals and monitor the GPS satellite errors. Any excessive satellite errors are indicated on a GPS integrity message broadcast by a master control center to GPS users. The GIC network must be capable of disseminating integrity data indicating the phases of flight possible and alerting the pilot to out of tolerance conditions within the required time-to-alarm.

Studies by the Working Group¹³ indicate that the GPS integrity may be determined relatively easily within the required time-to-alarm using a network of ground monitoring stations. The problem then remains of how to disseminate the data to the GPS users. The Integrity Working Group performed preliminary studies on both ground-based and satellite communication links for broadcasting the GIC message. This work is being continued by a GIC Working Group established by SC-159 to prepare a standard format for broadcasting GIC data.

2.3.1 Ground-Based Communication

Several methods of broadcasting GIC data using existing ground-based radio facilities were discussed by the Working Group¹⁴. The most attractive system considered was the use of Aeronautical Non Directional Beacons (NDBs) as a transmission medium. As stated in the U.S. Federal Radionavigation Plan²⁰, NDBs will remain a part of the radionavigation system well into the next century. It is expected that there will be 728 federal and 855 non-federal NDBs in the United States by the year 2000¹⁸. The NDB coverage extends over most of the United States and Canada. However, there are some areas (mostly the mountainous regions of the Western U.S. and Alaska) that do not receive service.

It would be possible to modify existing NDB stations to include the GIC data message modulated on the NDB signal without interfering with navigation operation¹⁸. The NDB range of operation for data transmission greatly exceeds the navigation range. However, the Working Group concluded that the existing NDB coverage would need to be extended to provide continuous coverage over CONUS suitable for a GIC integrity data link. The relatively low cost of operating and maintaining an NDB GIC link makes it an attractive alternative to a satellite based system.



Figure 4. Satellite GPS Integrity Network

2.3.2 Satellite Communication

A concept for a GPS Integrity network is shown in Figure 4. The GPS signals are monitored at ground-based stations linked through a ground communication network to a master control station. The master station uplinks the GIC data to geostationary satellites which then rebroadcast it to users in the area covered. To provide redundant integrity coverage over CONUS, two geostationary satellites are required. Five geostationary satellites would provide redundant world-wide coverage.

A variety of communication alternatives for the satellite broadcasts were discussed by the Working Group. Of concern to the group was the proliferation of systems (navigation, communication etc.) within the aircraft. Ideally, a GIC communication link would be sufficiently similar to the GPS signals that the GIC data could be received and interpreted directly by the GPS receiver itself.

Presently the GPS has 37 C/A codes reserved for use by the GPS satellites and ground transmitters. However, there are 1024 possible C/A codes which can be tracked by GPS receivers. If a geostationary satellite broadcast the GIC data modulated on the L1 frequency using an unassigned C/A code, the GPS receiver could track that signal and demodulate the integrity data internally. Because of the cross-correlation properties of the C/A codes, this integrity service would not interfere with the GPS navigation service.

In addition, it was suggested that the GPS-like signal broadcast by the geostationary satellite could also be used for navigation. In this case, not only is integrity ensured but the GPS satellite coverage is also enhanced. Providing two geostationary satellites over CONUS would supply sufficient redundancy to meet the FAA requirement for sole-means navigation, i.e. the service should not be interrupted by a single satellite failure. Sufficient spare bits are included in the existing GPS data format to provide a limited indication of the GPS satellites' integrity. It would also be possible to broadcast the navigation signal in phase quadrature with the integrity signal. This would provide a 50 Hz independent data channel for the GIC data.

A number of candidate satellite systems were discussed for broadcasting the GIC and GPS navigation messages. Two Block II GPS satellites were offered to the FAA by the GPS Program Manager, should the Department of Transportation pay for their launch costs. These satellites could be modified to broadcast the GIC data and to operate in geosynchronous orbit. A more inexpensive option considered was to piggy-back a GPS signal repeater on a geostationary satellite such as the GOES weather observation satellite¹⁷. The GPS-like signal would be generated by a ground-based master control station, accounting for the propagation delays in the uplink to the geostationary satellite. The signal would then be transmitted to the satellite where it would be mixed by the signal repeater to the L1 frequency and rebroadcast to users. This approach increases the complexity of the master control station but significantly simplifies the satellite payload.

Another approach considered was to use leased channels on Mobile-Service satellites or on commercial fixed service satellites¹⁸. This approach has merit in the short term as an inexpensive method of supplying GIC data. In fact, Inmarsat recently announced that they plan to offer a GPS integrity service to their users¹⁹. However, aircraft would be required to carry communication equipment in addition to the GPS navigation sets to use this type of GIC service. Certification problems may also arise with a navigation service (GPS) that is dependent on a communication service for integrity.

2.4 DIFFERENTIAL GPS

The use of differential GPS provides both integrity and improved GPS navigation accuracy. The differential GPS concept is similar to a GPS integrity monitoring network. The GPS satellite signals are monitored at surveyed locations to determine the satellite health and signal errors. The differential GPS message differs from GIC data in that error corrections are broadcast in addition to the satellite integrity data. This allows an out-of-tolerance signal to be still used for navigation once the error correction has been applied.

A standard differential GPS data format was prepared by RTCM Special Committee 104¹⁶ which allows common errors between the monitor station and the user to be eliminated. This improves the GPS navigation accuracy to 5 to 20 m 2dRMS as the SA errors can be completely removed from the navigation solution when the data is transmitted in a timely manner. Using differential GPS, it is therefore possible to meet the FRP goal of a 100 m alarm limit for non-precision approach. However, differential GPS does not increase the satellite coverage and so sole-means navigation would still be precluded by the FAA requirement for redundant satellite coverage.

3. CONCLUSIONS

3.1 SOLE-MEANS NAVIGATION

For GPS to be certifiable as a sole-means navigation system, the integrity of the navigation solution must be assured should system failures occur. Also, sufficient redundancy must be built into the system to allow navigation to continue in the event of a single satellite failure.

The navigation integrity can be provided through Receiver Autonomous Integrity Monitoring (RAIM) or by using a GPS Integrity Channel (GIC). With RAIM, the GPS receiver makes use of redundant navigation data for self-checking purposes. Redundant satellite data may be used or aiding data from other sensors on-board the aircraft such as baro-altitude, LORAN-C or an inertial navigation system. With a GIC, a network of ground monitoring stations continuously track the satellites and broadcast an integrity message indicating the health of the GPS satellites. These integrity monitoring techniques can meet all the integrity requirements and goals shown in Tables 1-1 and 1-2, except for the non-precision approach goal of 100 m.

The FAA coverage requirement, that the loss of a single satellite should not cause an interruption of service, is not met by the 18+3 GPS satellite constellation. Even without a satellite failure occurring, there are significant periods of time when unaided navigation is not possible with this constellation. Table 3-1 summarizes the methods of augmenting the GPS system that the Working Group considered were suitable for sole-means navigation with GPS.

If the satellite constellation were increased to 24 satellites, sufficient redundancy would be provided to allow navigation to continue in the event of a satellite failure and also to determine that a failure has occurred using RAIM. This would allow GPS to be certified as a sole-means navigation system for all phases of flight world-wide. Adding two geostationary satellites over the United States would provide sufficient redundancy to meet the FAA coverage requirement over CONUS. Integrity could also be supplied by broadcasting GIC data through the geostationary satellites. Five geostationary satellites could provide the same service world-wide.

The GPS system redundancy can also be increased by integrating the receiver with other sensors on board the aircraft. By including aiding data from a barometric altimeter, a GPS receiver can meet the integrity and coverage requirements for oceanic en-route navigation. An integrated GPS/INS can meet not only the oceanic en-route requirements (for which the INS is already certified) but also the tighter oceanic en-route goals and the domestic en-route integrity and coverage requirements. An integrated GPS/LORAN-C receiver can meet all the integrity requirements and goals except for the 100 m non-precision approach goal.

Table 3-1 Potential Sole-Means GPS Navigation Systems

GPS Navigation System	Phases of Flight	Oceanic En-Route		Domestic En-Route		Terminal Area		Non-Precision Approach	
		Req	Goal	Req	Goal	Req	Goal	Req	Goal
18 Satellites + 5 geostationary-GIC or 24 Satellite Constellation-RAIM		x	x	x	x	x	x	x	
18 Satellites + 2 geostationary over CONUS - GIC				x	x	x	x	x	
GPS/Baro-Altitude - RAIM		x							
GPS/INS - RAIM		x	x	x					
GPS/LORAN-C - RAIM				x	x	x	x	x	

3.1.1 Non-Precision Approach

Because of the planned level of the Selective Availability (SA) errors, the GPS navigation errors will exceed 100 m approximately 5% of the time, averaged globally over 24 hours. Since the SA errors change only slowly with time, the navigation error may stay outside the 100 m limit for a number of minutes. The Federal Radionavigation Plan (FRP)²⁰ established a 100 m alarm limit for future non-precision approach systems as a VOR located on the runway supplies 100 m accuracy (95%) at 0.7 nmi from the airport. This means that during the approach, the VOR can be expected to be within this accuracy 95% of the time. However, with GPS 1° the SA errors cause the navigation accuracy to exceed 100 m at the beginning of the flight, the same navigation error will apply for the next few minutes. In the worst case condition, the 100 m alarm limit could be exceeded 100% of the time throughout the approach. Essentially, the GPS 100 m 95% accuracy limit set by the SA errors means that only 95% of the time can an approach be made using GPS that will compare in accuracy to a VOR non-precision approach.

The Working Group concluded that the SA errors would preclude GPS meeting the non-precision approach standards established in the FRP, unless the SA errors were reduced or GPS was operated in the differential mode.

3.2 SUPPLEMENTAL NAVIGATION

In the near term, the most promising application for GPS is as a supplemental navigation system. During the constellation build-up phase, sole-means navigation will not be possible due to insufficient satellite coverage. However, as a supplemental navigation system, the excellent navigation accuracy provided by GPS may be used whenever sufficient satellites are in view to allow navigation.

The GPS navigation integrity may be ensured either through cross-checking with the sole-means navigation system or through making use of the integrity monitoring techniques suggested for sole-means GPS navigation. Table 3-2 lists the supplemental GPS applications that were considered by the Integrity Working Group.

Table 3-2 Supplemental GPS Navigation Systems

GPS/Multi-Sensor FMCS	GPS/VOR/DME-RNAV
GPS/Omega	GPS/INS
GPS/LORAN-C	

3.3 RECOMMENDATIONS

The Integrity Working Group made the following recommendations to the SC-159 committee in their final report:

1. Expanded GPS Constellation

As recommended by the Working Group, a letter was drafted by the RTCA to the Department of Defense urging that the GPS constellation be expanded to 24 satellites. The 24 satellite constellation would provide continuous coverage and sufficient integrity for civil navigation world-wide.

2. Reduction in the Effect of Selective Availability

Analyses by the Working Group showed that the presently specified level of selective Availability (SA) is unacceptable to meet the non-precision alarm limit goal of 100 m. The Working Group recommended that a joint committee be formed with DoD personnel to (a) determine the appropriate alarm limit required for non-precision approach, (b) determine the SA level and bounds consistent with this requirement, and (c) study the effects of the planned SA errors on GPS integrity. The acceptability of a reduced level of SA to the DoD should also be assessed.

3. GPS Integrity Channel (GIC) Working Group

As recommended by the Integrity Working Group, on September 9, 1987, SC-159 formed a GPS Integrity Channel (GIC) Working Group. This group was tasked to prepare a standard GPS Integrity Channel (GIC) data format for communicating the integrity of the GPS navigation solution. The group will consider civil GPS user requirements, monitoring/communication system requirements, GIC data message formats and suitable GIC communication media.

4. Topics for Further Study

The Working Group recommended that studies should continue on Receiver Autonomous Integrity Monitoring techniques, GPS/LORAN-C integration techniques and the use of Baro-Altitude aiding for integrity monitoring.

4. REFERENCES

1. SC-159 Integrity Working Group, Radio Technical Commission for Aeronautics, "Report of the SC-159 Integrity Working Group", May 1987, RTCA Paper No. 220-87/SC159-95
2. Kalafus R. and Chin G., "Measures of Accuracy in the Navstar/GPS: 2dRMS vs CEP", ION National Technical Meeting Proceedings, Long Beach, Jan 1986

3. Kruh P., "The Navstar Global Positioning System 6-Plane, 18-Satellite Constellation", Nat. Telecommunications Conf. Proceedings, New Orleans, Nov 1981
4. Brown R.G., Radio Technical Commission for Aeronautics, "Selective Availability Simulations", RTCA Paper No. 19-87/SC159-73
5. Francisco S., Radio Technical Commission for Aeronautics, "Operational Control Segment of the Global Positioning System", RTCA Paper No. 117-86/SC159-16
6. Brown A. and Jessop R., Radio Technical Commission for Aeronautics, "Receiver Autonomous Integrity Monitoring using the 24-Satellite GPS Constellation", RTCA Paper No. 279-87/SC159-102
7. Brown R.G., Radio Technical Commission for Aeronautics, "GPS Failure Detection by Autonomous Means within the Cockpit", RTCA Paper No. 247-86/SC159-26
8. Lee Y., Radio Technical Commission for Aeronautics, "Analysis of Range and Position Comparison Methods as a Means to Provide GPS Integrity in the User Receiver", RTCA Paper No. 248-86/SC159-27
9. Brown R.G., Radio Technical Commission for Aeronautics, "Self-Contained GPS Failure Detection", RTCA Paper No. 318-86/SC159-39
10. Brown A., Radio Technical Commission for Aeronautics, "Integrity Monitoring of GPS using a Barometric Altimeter", RTCA Paper No. 405-87/SC159-117
11. Kalafus R., Radio Technical Commission for Aeronautics, "Durations of Periods with Less than Five Satellites in View", RTCA Paper No. 488-86/SC159-59
12. Brennen M., Radio Technical Commission for Aeronautics, "GPS Monitoring using IRS or INS", RTCA Paper No. 461-86/SC159-50
13. Jorgensen P., Radio Technical Commission for Aeronautics, "A Suggested Approach for Integrity Monitoring Stations", RTCA Paper No. 18-87/SC159-72
14. White F., Radio Technical Commission for Aeronautics, "Communications Methods for Disseminating GPS Integrity Data to Aircraft using Planned Capability", RTCA Paper No. 303-86/SC159-33
15. Scull D., Radio Technical Commission for Aeronautics, "Communications Link for Differential GPS Monitor Stations", RTCA Paper No. 457-86/SC159-46
16. Special Committee 104, Radio Technical Commission for Maritime Services, "Recommendations of Special Committee 104: Differential Navstar/GPS Service", Nov 1985
17. Ford Aerospace and STI, Radio Technical Commission for Aeronautics, "Implementation of a GPS Type Payload on a Geostationary Satellite", RTCA Paper No. 246-86/SC159-25
18. Braff R., Radio Technical Commission for Aeronautics, "GPS Signal Integrity Briefing for RTCA SC-159", RTCA Paper No. 114-86/SC159-14
19. "Inmarsat moves to provide satellite navigation services", Aeronautical Satellite News, Sept 1986, p1
20. Department of Transportation, "Federal Radionavigation Plan (FRP)", DOT-TSC-RSPA 84.8

INTEGRATION OF GPS AND STRAPDOWN INERTIAL SUBSYSTEMS INTO A SINGLE UNIT

by

David Buechler — Advanced Systems Manager
Michael Foss — GPS Development Manager
Precision Products Division
Northrop Corporation
100 Morse Street
Norwood, MA 02062 (617) 762-5300
United States

SUMMARY

Most GPS receivers are designed for stand-alone operation. Single-channel, slow-sequencing receivers are recognized to be lowest cost, but lack dynamic capability, while multiplexed and multichannel designs allow operation during acceleration, but add cost, weight and power. It is also well recognized that a marriage of GPS data with strapdown inertial data enhances the quality of both, and many people are looking at integration of available GPS receivers with various available inertial systems. However, direct design at the outset of a tightly integrated GPS and strapdown INS allows optimization of both for performance and cost. Using this approach, a properly designed slow-sequencing, single-channel receiver, married to a low-cost strapdown inertial unit, provides satellite tracking during 10-g acceleration, very high jamming suppression, improved strapdown inertial outputs, and improved GPS navigation accuracy. Packaging this GPS/I in a single unit reduces software and hardware redundancies, and results in a very low-cost design.

INTRODUCTION

Reference [1] is a detailed comparison of multichannel vs single-channel receivers. Most drawbacks of the lowest-cost, single-channel, slow-sequencing design are related to its lack of dynamic capability. Aiding the tracking loops with high-frequency velocity and position from a strapdown Inertial Measurement Unit (IMU) eliminates this disadvantage while simultaneously providing increased jamming suppression. The GPS/I unit also provides the benefits of inertial attitude, attitude rate, acceleration and extrapolation of position and velocity whenever the GPS satellites are lost due to bad GDOP, shadowing, jamming, or receiver failure.

Even a moderately accurate IMU can become a good navigator if heading error is removed, and the gyro and accelerometer turn-on errors calibrated. The high accuracy position and velocity from the GPS data provides these calibrations. Conversely, obtaining first fix on four GPS satellites is difficult on a dynamic vehicle. Short-term navigation from the IMU is adequate to obtain a first fix in less than two minutes in the integrated design, even with acceleration and maneuvers involved with takeoff or air launch of a missile or RPV using this GPS/I unit.

Mutually aided GPS and IMU subsystems have been discussed in previous papers [Ref. 2 and 3], but have so far been designed around existing equipment. During 1983 and 1984, PPD teamed with Rockwell Collins and Boeing Aerospace to test such a system, the results of which were presented at the 1985 ION meeting in San Diego and later in [Ref. 4].

By using available high-frequency digital technology and LSI, minimal analog hardware is required and packaging of the receiver as an integral part of the INS is possible. The single clock crystal, which runs all equipment, may be combined with the inertial cluster to compensate acceleration sensitivity and provide vibration isolation. Cascaded Kalman filters and time correlation of measurements are eliminated. A single I/O and power supply reduce cabling and weight compared to multiple units.

During development of the P-code receiver and GPS/I system, each phase was verified using simulations and tests. Phases in this recent activity include: GPS/I architecture selection and simulation verification of performance; verifying the performance of the GPS receiver and integrated GPS/I breadboards against the original design goals and the simulations through antijamming and dynamic testing; design, manufacture, and testing of custom LSI chips for the digital portions of the receiver; integration of the production prototype with CRPA antenna electronics; and van and flight testing of a GPS/I Navigation System (GPS/INS).

The GPS/I unit can be divided into the following five functional blocks: IMU Sensor Cluster, RF front end, digital signal processing and frequency synthesizer, GPS and navigation computers, and interfaces.

100

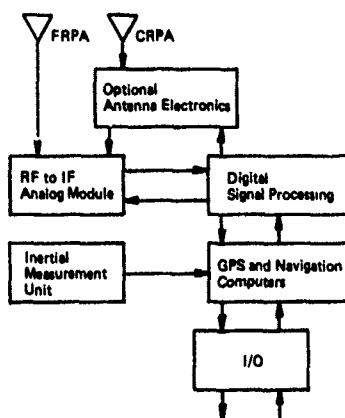


Figure 1. GPS/I Functional Partitioning

Figure 2 shows the GPS/I unit packaged into the Navy's Standard Attitude Heading Reference System enclosure, which also contains a power supply, fan, connectors, and EMI section. The box is 7 x 7 x 11 in. and weighs less than 24 pounds. The inertial sensor cluster uses a pair of DTGs and three linear accelerometers.

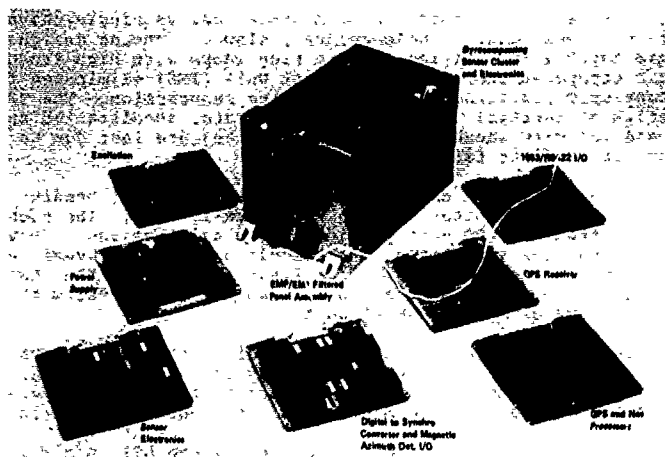


Figure 2. Proposed GPS/I Unit Packaged into SAHRS

GPS RECEIVER SUBSYSTEM

The GPS receiver is configured to operate with the receiver's single channel dedicated to each satellite, (SV), for a period of one second. An exception to the sequential mode is used during initial SV acquisition to provide a time-to-first-fix (TTFF) of less than two minutes whether stationary or on a moving vehicle. The first satellite is acquired through rapid code slew depending upon uncertainty in knowledge of initial position, velocity, and time of week. The ephemeris is taken and time is synchronized to the C/A code epoch. During ephemeris collection from the last three of the required four satellites, the receiver software is configured to operate the hardware in a fast data collection mode, providing simultaneous data collection from all three. This feature requires no increase in hardware complexity, since it is accomplished by setting up the C/A code generator at the beginning of each dwell by rapid slewing. The maximum required slew time regardless of range differences between satellites is 0.5 msec. This fast data collection mode is not used for P-code operation, avoiding increased P-code generator complexity and inherent loss of anti-jam tolerance.

RF Front-End. The RF front-end schematic is shown in figure 3. Stages 1 through 8, are identified as the receiver RF electronics. Stage 1, which provides low-noise amplification of the L1 and L2 signals, is a discrete design that is impedance matched to 50 ohms by using microstrip transmission lines and a limited amount of resistive feedback. The amplifier immediately following the LNA is a discrete device with internal resistive feedback to establish the requisite gain and device port impedance levels.

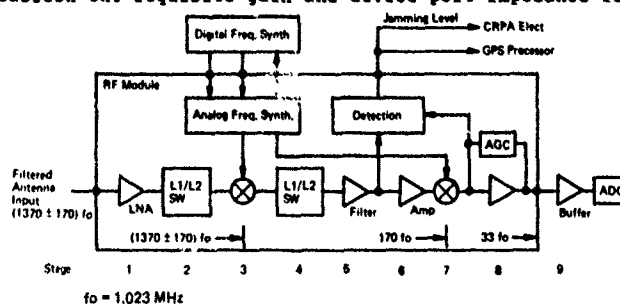


Figure 3. RF Front-End Functions

Stages 2 and 4 provide L1 or L2 selection, consisting of two SPDT switches and two bandpass filters. The two filters are centered at L1 and L2. The mixer (stage 3) down converts the L1 or L2 signals to the first IF frequency of 173.91 MHz, ($170 f_0$ where $f_0 = 1.023$ MHz). The input and output ports of the mixer have very low reflection coefficients with port VSWRs of typically less than 1.5 to 1. The good impedance match to 50 ohms is a result of the action of the quadrature couplers.

The first IF stage (Stages 5-6) consists primarily of gain and the IF filter. This filter establishes a receiver RF bandwidth to optimize signal-to-noise ratio. The second mixer downconverts the first IF signal to the second IF frequency of 33.759 MHz, ($33 f_0$). The second IF stage (Stage 8) provides further amplification and contains the AGC amplifiers.

Careful consideration must be given to the receiver linearity and AGC operation when dealing with very large J/S ratios. The weak GPS signal superimposed on the strong jammer must not be suppressed. The receiver must be linear and the AGC must stabilize the signal envelope input to the two-bit AD converter (ADC).

The post-mixer IF amplifier is a low-noise, discrete transistor design. All the remaining IF amplifiers are low-cost silicon monolithic integrated circuits.

Digital Signal Processing. The key components of this block comprise a pair of identical adaptive threshold, two-bit analog/digital converters (ADCs), integrators, P-code and C/A code generators, a code phase controller, and numerically controlled oscillator (NCO) to control the carrier, and its associated sampler control circuitry. These functions are shown in figure 4.

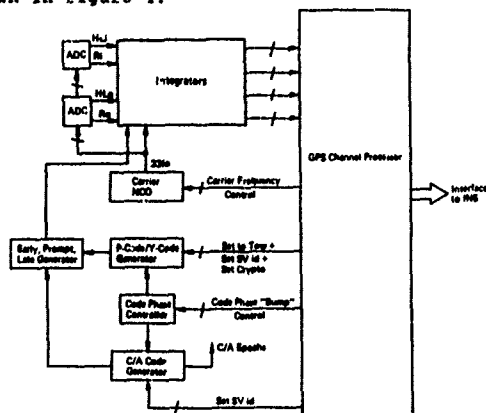


Figure 4. Digital Signal Processing Functions

The output from the second IF, nominally at $33 f_0$, is applied in parallel to the two ADCs, clocked in phase quadrature at a nominal rate of $33 f_0$. Each ADC output is a two-bit quantized baseband signal at a rate of $33 f_0$ samples/second.

The ADCs maximize performance over a wide range of jamming conditions as described in [Ref. 5 and 6]. The ADC constitutes the nonlinearity in a nonlinear matched filter, matched to the incoming (GPS) signal waveform, providing the maximum possible output SNR. Nonlinear filters can outperform linear filters when the signal is accompanied by non-Gaussian type interference, while providing minimal degradation over the optimum linear filter in Gaussian-type interference. [Ref. 5]

Each ADC quantizes to two-bit accuracy, producing outputs HL (Hard-Limit) and R (Ratio) defined in figure 5. A value of 4 for the weight, R applied to samples which occur above the upper threshold, Δ , represents the best performance compromise between Gaussian and non-Gaussian types of interference. [Ref. 6] The upper threshold level, Δ , is controlled by a threshold-adjust circuit, consisting of a control loop which monitors the R output and moves the threshold to maintain a predetermined percentage of samples (X percent) at the R=4 level. Correct selection of X determines the anti-jamming performance of the ADC and much simulation work has been completed to determine the optimum choice for X.

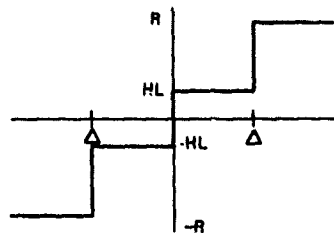


Figure 5. Two-Bit A/D Transfer Function

A primary objective in the receiver design was to sample and quantize the signal directly at the IF to eliminate the need for analog demodulation (conversion to baseband), followed by baseband sampling. The principles of passband sampling are well documented in the literature [Ref. 7], but its widespread adoption has been limited until recently by the availability of economical high-speed digital dividers.

Baseband sample trains (comprising two pairs of signals, in phase and quadrature) are mixed with Early, Prompt, and Late (E, P, L) sequences from the receiver's P-code or C/A code generator. Four sets of despread signals are thus formed and applied to the inputs of the integration channels. From these signals, the code and carrier phase error can be calculated. This type of correlation reduces the complexity and number of registers required to determine the code and carrier tracking errors. Despreading the GPS signals in the digital domain eliminates the need for a T-code signal, required for good anti-jamming tolerance in receivers using analog despreading. [Ref. 8]

Each integration channel accumulates its input samples synchronized with the local C/A code millisecond epoch. Alignment of the millisecond integration period with transitions in the 50-baud navigation data stream is thereby ensured. The values are read into the GPS processor and processed to recover carrier and code tracking information, pseudo-range and range rate, signal strength/quality data, and the 50-baud navigation data. Samples from the integration channels are processed by the carrier and code tracking algorithms.

The C/A and P-code generators produce early, prompt, and late versions of the selected code and support double-speed code slewing. The code-phase controller receives commands from the GPS processor instructing it to adjust the code phase in fractions of a code chip to facilitate tracking of the satellite code signals. The clocks supplied to the C/A and P-code generators by the code-phase-controller are slaved together; the C/A code generator is always kept running, even when tracking P-code.

The composite signal loss for the above design is 1.8 dB. This compares favorably with losses typically in the range 2 to 3 dB for analog designs. All losses have been calculated for Gaussian noise (the worst possible noise type for loss budgeting).

Frequency Synthesizer.

The frequency synthesizer generates local oscillator (LO) signals to drive the two downconversion mixers in the RF front-end.

Factors Affecting the Choice of Sampling Rate.

1. The sampling rate should be high enough to provide minimal performance degradation in the digital matched filtering process. In practice, this requires a minimum sampling rate exceeding two to three times the P-code chip rate.

2. The sampling rate should be asynchronous to the P-code chip rate to provide a smooth E-L codetracking discriminator characteristic. For C/A code, a sampling rate either very much higher than, or asynchronous to, the chip rate is satisfactory.
3. The sampling rate should be kept low to be compatible with use of power and size-efficient LSI digital signal processing technologies.
4. ADC design is simplified by minimizing the sample rate.

Taking these factors into account, the IF and sampling frequencies were each selected with the nominal value of 33 fo (33.759 MHz).

IMU SUBSYSTEM

A representative IMU selected for the GPS/I unit is derived from the Navy's Standard Attitude Heading Reference System (SAHRS). The IMU contains two dynamically tuned dual-axis gyros and three force rebalanced accelerometers. The electronics provide all the control and interface functions needed to make this a line replaceable module within the GPS/I.

Multiplexed rebalance capture loops, used on all PPD systems, reduce the parts count. The gyro rotor deflection angle signals from the gyro pickoffs are demodulated, filtered, frequency compensated, and multiplexed into an analog-to-digital converter. The output is processed by the CPU and sets the switching point of the pulse-width-modulated signals that drive the torquer which holds the gyro at null.

The gyro channels have dual rate range capability to accommodate the dynamic inputs in a power-efficient manner, up to 200 degrees per second in the high range.

The accelerometers are servo force rebalance instruments, providing a voltage output proportional to acceleration. This signal is integrated, compensated, and multiplexed to the A/D for conversion to a delta velocity digital word. The integrator is nulled using the same PWM rebalance technique as the gyros.

An eight-channel analog multiplexer sequentially samples the seven inertial outputs and uses the eighth channel to sample sensor temperature signals on a submultiplex basis. Each inertial channel is sampled at a 1-KHz rate by a microprocessor, integral to the sensor cluster. Calibration constants for the IMU are determined during acceptance testing from a series of static and dynamic tests at different temperatures, and programmed into the PROM. These coefficients are up-loaded to the navigation computer during system initialization. The calibration data stored in the PROM consists of:

1. Gyro bias, scale factor, scale factor asymmetry/nonlinearity, and mass unbalance with associated temperature coefficients per sensor channel and per rate range (as required).
2. Accelerometer bias and scale factor with temperature coefficients per sensor channels.
3. Inertial sensor misalignment matrix.

The navigation computer executes data compensation algorithms using the PROM constants to effectively eliminate deterministic errors from the incremental angle and velocity IMU data.

All required voltages are provided by the internal power supply. A 16.368-MHz (16 fo) signal from the GPS clock generates all IMU timing and control signals to synchronize the data signal generation between the IMU and the GPS subsystems. The timing/control function is contained in a custom integrated circuit used in SAHRS. It also contains the BIT functions and two programmable timers that provide the dynamically tuned gyro spin motor frequencies.

COMPUTER SUBSYSTEM

The GPS and NAV computers use the same MIL-STD-1750A architecture. Each processor is coupled with its own 64K words of memory and input/output devices through a demultiplexed address/data bus. This approach readily accommodates a range of device interfaces. The processors are data coupled through a RS-422 serial link and time synchronized from a common clock in the GPS Receiver.

Nonvolatile memory (NVM) for critical data storage during power off is provided as required by the system application and need only be available to the NAV processor. Satellite almanac and other available initialization data are supplied to the GPS processor from NVM on power-up.

The required software functions have been partitioned from both functional and processor-loading aspects. The GPS processor provides the GPS receiver control, satellite data-gathering functions, and satellite range and range-rate measurements.

The navigation processor implements the strapdown navigation function, Kalman processing including both transfer alignment and satellite measurements, the host vehicle interface, system state data (position, velocity, attitude, and GPS clock offset) to aid the GPS receiver processor signal acquisition and tracking. Each processor uses a priority-based multitasking executive.

GPS PROCESSOR SOFTWARE

The GPS receiver software has been designed to take full advantage of the facilities offered by the digital signal processing section of the receiver hardware and the underlying philosophy of tight GPS/I integration. The primary aim of the receiver software is to reach a steady-state sequencing mode of operation, generating measurements to allow the navigation filter to estimate GPS clock and navigation sensor errors and bound the navigation solution.

Satellite Selection - The satellite selection algorithms are implemented using mathematical set theory, allowing easy manipulation and formal proof techniques. Processing time is minimized by using the volume of the pentahedron formed by four satellites and the host vehicle as a first-order approximation to geometric dilution of position (GDOP).

Tracking parameters are selected using a nondeterministic, decision-making, finite-state machine (FSM). This FSM consists of seven prioritized states shown in figure 6.

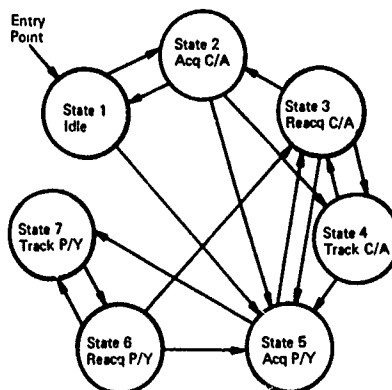


Figure 6. Acquisition Management Finite State Machine State Transition Diagram

Each state consists of a set of substates. Transition from one state to another takes place whenever predefined entry conditions for the destination state are satisfied by the current environmental descriptors. On entry to a destination state, these descriptors are further analyzed for acceptance by one of the state's substates. The substate descriptors used as inputs to the decision-making process include:

- Receiver configuration (code generation facilities, encryption capabilities)
- Satellite health, elevation angle and anti-spoof status
- Time alignment status
- Availability and values of J/S and CNO measurement.
- Availability and age of satellite ephemeris and almanac data
- Host vehicle position and velocity error estimates (from navigation filter)
- Availability and age of pseudorange and delta-range measurements

Tracking parameters are predefined for each of the substates. Therefore, whenever an acceptable state and substate pair is found, an associated set of tracking parameters is made available to the prepositioning and tracking loop software for initiation of the next dwell.

Satellite Acquisition. Based on available data, the receiver will determine the optimal method (sky-search, C/A search, direct P acquisition) to acquire satellites.

If the receiver does not have satellite almanac data and approximate time, it will execute an open-sky search (a search of the whole C/A epoch and all possible Doppler frequency offsets) for each of the 18 satellites until one is acquired. Almanac data will then be decoded from the 50-Hz data stream for all satellites. Satellite almanac

information taken in this manner may take up to 12.5 minutes to collect, but is then stored permanently for later use. This data will be valid for a period of about 6 months.

If almanac data is available from a prior collection, or has been supplied from a host vehicle at turn-on, a C/A acquisition is initiated, which requires a knowledge of host vehicle (HV) position and velocity to within 100 km and 70 m/s, time to within 10 minutes, and the almanac for a C/A acquisition.

The timing of receiver functions with respect to the incoming satellite 50-Hz data subframes is shown in figure 7. This demonstrates that a TTFF (using the P-code with ionospheric corrections) is achievable in 108 seconds, worst case, nominally only 82 seconds.

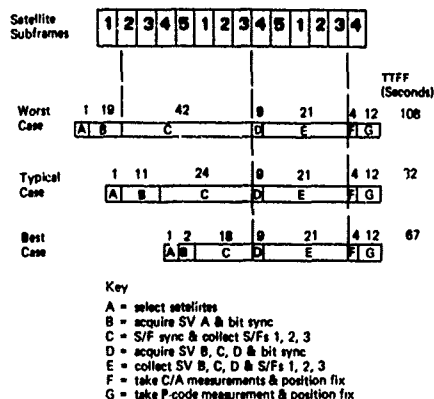


Figure 7. Fast Data Acquisition Time to First Fix (TTFF)

A sequential detection algorithm, implemented in software, is used for locating a C/A signal.[9] This is initialized using all available information regarding satellite Doppler and code phase. For acquisition of the first satellite prior to clock setting, a current almanac and information from the IMU are used to predict the satellite's Doppler cell, but the code is searched over its entire uncertainty region. For subsequent satellites following clock setting, the search is initialized for both Doppler cell and anticipated code-phase, minimizing search time for these satellites.

The detector algorithm measures the power level in the current search cell during eight 1-millisecond samples, after which the result is compared with an adaptive threshold (set in accordance with current jamming level). If, at the end of any dwell, the power is below the threshold, the signal is assumed absent and the search is moved to the next search cell. Code cells are searched at a relatively high rate of 60 chips/second, facilitated by a novel parallel search technique, giving a 50-percent increase in search rate compared with conventional sequential techniques.

After locating the signal, a series of C/A centering procedures is followed which require about two seconds.

Carrier Tracking. During state 5 tracking of P-code (while sequencing), twenty 1-millisecond samples from each of the integration channels are summed within the processor (over one period of a nav data cycle which runs at 50 Hz) to form five 20-msec sample values; i.e., the predetection interval (PDI) is 20 msec. Because of the tight integration of IMU data, the PDI is maintained at 20 msec throughout the 1-second dwell on any given satellite during state 5 tracking. There is no need for reduced PDI, which degrades anti-jamming margin, during the early portion of a dwell, as with some designs not specifically intended for incorporation of inertial inputs. Lower values of PDI are used during initial C/A code acquisition (1 msec), and during C/A code fast data collection (6 msec).

Code Tracking. For state 5 GPS operation, a noncoherent first-order E-L code tracking algorithm is used, which has a 3-dB advantage in SNR compared with a tau dither tracker. The code loop is continuously frequency aided from the carrier loop or the IMU, so that a narrow-band loop may be used. The use of a noncoherent, rather than coherent algorithm, accommodates high dynamics applications where the carrier phase error is not always a fraction of a carrier cycle.

The receiver algorithm operates in state 3 similar to that for state 5, except that the code error is not applied to a loop filter, but is averaged over the period of a dwell, linearized, and passed to the navigation processor to be used in the Kalman filter. The navigation computer returns the information to set the code and carrier during the next dwell.

Sequential Resynchronization. A 1-second satellite dwell structure is used during sequencing in state 5. This dwell is divided into the subdwells described in the following sections. The predetection interval is maintained at 20 msec throughout the dwell, thus ensuring maximum anti-jamming capability.

Subdwell 1. This 20-msec period is used for setting up the CRPA electronics if used for the new satellite. Attitude information is conveyed to the antenna to facilitate beam pointing. The P-code generator, encryption circuits, and sampler are also initialized during this period. Code phase is set in accordance with prepositioning data obtained from the navigation computer.

Subdwell 2. This subdwell is used for initial P-code centering, using a first-order noncoherent code loop of 2.5-Hz bandwidth. The carrier loop is not enabled here. The carrier frequency and frequency aiding to the code loop are therefore slaved to the IMU (velocity and acceleration) during this period.

Subdwell 3. The code loop will now have acquired, allowing the arc tangent carrier tracking loop to be enabled (the loss in its threshold at the beginning of this subdwell is <0.25 dB). The code loop is run with a 1-Hz bandwidth and the carrier loop with an 8 Hz bandwidth. The carrier loop is acquisition aided with both velocity and acceleration from the IMU.

Subdwell 4. This period is a continuation of subdwell 3, but with code loop bandwidth reduced to 0.5 Hz.

Subdwell 5. Collection of delta range information from the carrier loop is undertaken during this period. Range data is collected from the code loop throughout dwells 2 to 5.

During dwells 3 to 5, the 50-baud navigation data is recovered, using differentially coherent detection. This facilitates data recovery over a larger portion of the dwell and with far higher dynamics than possible with coherent data demodulation.

On initial entry to the sequencing mode, C/A measurements are available for the four satellites comprising the initial constellation. Acquisition of the P-code is achieved using these measurements together with the receiver's estimate of GPS time. Switching between satellites is effected by a combination of a preload of the P-code generator and either accelerated or decelerated generation of the P-code for up to 1 millisecond. When the receiver has collected measurements using the P-code at the L1 frequency for each satellite, a set of measurements at the L2 frequency is taken for the purpose of ionospheric correction.

During the first two to three minutes of sequencing, measurements of pseudorange and delta-range are used by the Kalman filter to calibrate the GPS receiver clock, sensor, and navigation errors. When the Kalman filter has sufficiently reduced the system error estimate, the receiver initiates collection of the 50-Hz satellite navigation data for all other visible satellites. The steady-state, 1-Hz sequencing is interrupted periodically by two submodes, as follows:

1. Acquisition of the L2 frequency for ionospheric correction measurements once each minute for a period of 2 seconds. The single-channel sequencing receiver cannot take measurements at both L1 and L2 frequencies simultaneously, and during the lag between measurements, the host vehicle may move. A 2-second dwell is used, within which a measurement at the L2 frequency is sandwiched by two measurements at the L1 frequency. An L1/L2 measurement is then derived (aided by position measurement from the Kalman filter) by interpolation of the two outer L1 measurements to the time of validity of the L2 measurement.
2. Whenever the receiver database indicates out-of-date or uncollected ephemeris data for any visible satellite, 6-second dwells are used to collect satellite navigation data. The frequency of this operation thus depends on the number of visible satellites, but will nominally occur in bursts of three, for collection of each of the first three subframes in the satellite navigation message data, with a frequency of once every 4 hours.

For direct P-code acquisition, position and velocity to 1 km and 2 meters per second, time within 1 microsecond and ephemeris data is required. Figure 8 shows that a TTFF direct P-code acquisition can be accomplished in 10 seconds.

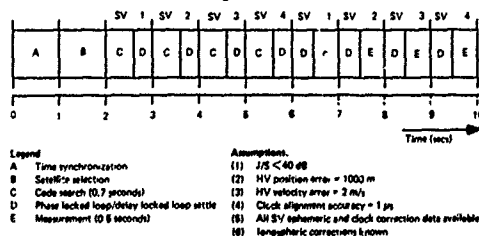


Figure 8. Direct P-Code Acquisition

NAVIGATION PROCESSOR SOFTWARE

The navigation software provides the functional capability summarized by the following major application tasks.

1. Compensate IMU gyro and accelerometer data using calibrated error corrections.
2. Initialize and maintain (100 Hz) the estimate of host vehicle position, velocity, and attitude.
3. Compute measurements (range, range-rate) performed by the measurement processing function using the satellite data provided by the GPS receiver processor and the IMU navigation state estimate.
4. Estimate system errors (position, velocity, attitude, instrument bias, receiver clock) provided by the Kalman filter using receiver measurement data at a 1-Hz rate.
5. Provide receiver aiding data (host vehicle position, velocity, attitude) at a 10-Hz rate.
6. Provide interface to the host vehicle data bus. This function controls the formatting and transmission of required output data in addition to reception and deformatting of commands received from the host vehicle.
7. Continuously monitor the system and processor health and report faults as detected.

Navigation Equations

The navigation equations continuously estimate the host vehicle position, velocity, and attitude by processing the IMU gyro and accelerometer incremental angle and velocity data in a fourth-order quaternion S/D algorithm. The navigation computations include compensation for calibrated instrument errors.

The data flow for navigation processing (figure 9), shows the raw instrument data accumulated at 100 Hz and adjusted for calibrated instrument errors and known correction terms. The vehicle attitude is updated at a 100-Hz rate using gyro-measured body rates. Vehicle attitude is used to resolve the body-fixed accelerometer outputs into the wander azimuth navigation frame where coriolis corrections are applied and ground speed is calculated. The ground speed vector is used to compute the vehicle navigation rate relative to earth-fixed axes. The vehicle geodetic latitude and longitude are then computed using the appropriate elements of this matrix.

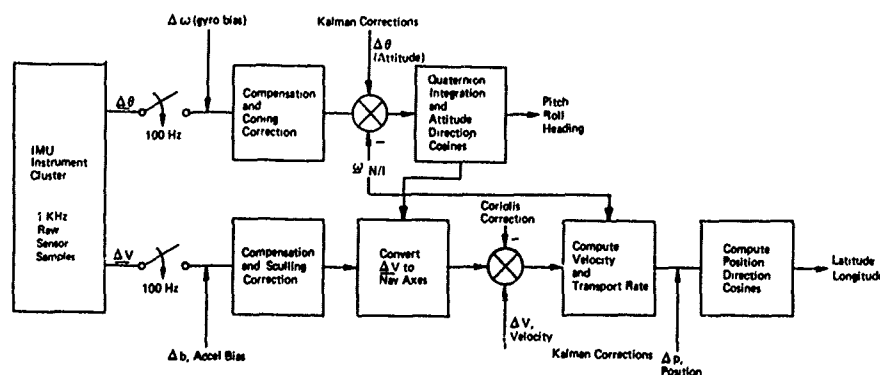


Figure 9. Navigation Equations Data Flow Diagram

The vertical channel mechanization for this system is a GPS-aided mechanization shown in figure 10.

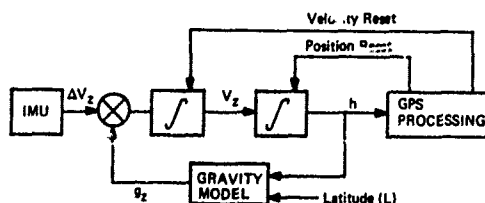


Figure 10. Vertical Channel Navigation

Kalman Filter

The Kalman filter mechanization chosen for this system is a 17-state upper diagonal filter. This filter has computational advantages over other approaches and has been successfully demonstrated to provide high navigation performance. The filter includes:

Error States	No. of States
Latitude, longitude, altitude	3
North, east, down velocity	3
Roll, pitch, heading	3
Gyro drift bias	3
Accelerometer bias	3
Clock bias	1
Clock drift	1

HOST VEHICLE I/O

Data transfers with the host vehicle are controlled by the navigation processor and are implemented via either a 1553 I/O or an RS422 interface. The I/O controller is memory mapped to the navigation processor. Output data is transmitted synchronously with the strapdown computations to minimize any output data latency. Typical output parameters and their associated rates include:

Parameter	Rate (Hz)	Number of Words
Attitude (roll, pitch, yaw)	100	3
Body rates (roll, pitch, yaw)	100	3
Acceleration	100	3
Latitude, longitude	25	2
Altitude (AGL or MSL)	25	1
Velocity (N, E, D)	25	3
Velocity (ECEF)	25	3
Position (ECEF)	25	3
Direction cosine (ECEF to body)	25	9
Navigation quality	25	1
Unit up vector (ECEF)	25	3
Gravity	25	1

The host vehicle will supply initialization data to the navigation processor including position, velocity, satellite almanac data, GPS time of week, and the signal encryption key.

GPS/I EVALUATION

The architecture described above has been fully evaluated and tested through simulation, laboratory tests and dynamic van and flight tests.

Simulation to Verify Performance

A comprehensive simulation program has modelled the selected receiver architecture. A primary statistic generated by this model is the SNR produced at the output of the integrators. Sensitivities of this statistic to the receiver front end, the ADCs, the sampling strategy, and the predetection interval have been investigated for a range of interference scenarios. Results from this simulation have been confirmed by laboratory measurements on the prototype P-code receiver.

The simulation program accurately models the receiver architecture and provides the following features:

1. Signal generation and filtering. Signals are generated to mimic the BPSK GPS P or C/A code signals, including accompanying Doppler offsets and 50-baud navigation data. Thermal noise is added to the signal to provide a specified CNO as measured at the receiver antenna input. Both signal and interference are applied to a digital filter algorithm, whose shape has been selected as characteristic of that used in the receiver front end. The effect of the AGC circuit on the composite input signals is included.
2. Quantization. The two-bit ADC is accurately modelled, down to the effects of imperfections such as threshold asymmetries, DC biases, and hysteresis effects. The adaptive threshold is controlled using a representative feedback control algorithm.
3. Sampling. The receiver's sampler is accurately represented, with the effects of sampling jitter, caused by finite clocking rates and resolution of the PPO circuits, and quadrature timing inaccuracies between the in-phase and quadrature clocks (quadrature errors) being included.
4. Jammer Types. The ability to represent CW, wideband noise, AM, FM and pulsed jammers of definable J/S value is included.

CW Jamming Simulation. A wide range of jamming strategies was evaluated using the model, with the performance of the ADC highly satisfactory. These results have also been confirmed experimentally.

Threshold Adjust Time-Constant Selection. Correct selection of the natural frequency, f_l , of the ADC threshold adjust circuit is important.

The sampled output generated by each ADC will, for the case of CW jamming, be a sinusoid at the difference frequency, f_d , between the jammer and the GPS carrier. When f_d is large relative to f_l , the adaptive threshold will be substantially constant and behavior of the circuit will be as seen in preceding sections. When f_d is much less than f_l , the threshold will track the jammer sinusoid and the jammer will be almost entirely rejected. However, when the jammer offset is close to f_l , the threshold will attempt to track the jammer, but with a phase lag and attenuation that will disturb the normal action of the ADC. To circumvent this problem, f_l should be selected at a small enough value to minimize the chance of the jammer offset being close to f_l for any significant time period. A value of $f_l = 50\text{Hz}$ satisfies this requirement.

Simulations have been performed to assess the effect of a jammer offset, f_d , made equal to f_l . It has, however, proven practically impossible to achieve synchronism between the jammer and satellite signals long enough to unlock the receiver tracking loops, even when the J/S is close to the receiver thresholds. The task would be even more difficult with high vehicle dynamics. In view of these considerations, the jammer will achieve no benefit in trying to set $f_l = f_d$.

Laboratory Testing

A complete prototype version of the P-code receiver has been constructed, consisting of nine wirewrap cards for the digital processing elements, a screened module housing the RF front end, and an HP1000 computer used as the GPS processor. This breadboard was used to prove the principles of the design, and associated software algorithms, prior to fabrication of custom LSI circuits.

The receiver prototype has been used to determine performance with a variety of jamming signals, while tracking either simulated or real satellites. All results show good correlation between measured and predicted receiver thresholds for continuous (single-satellite) state 5 operation. Indistinguishable results have been obtained while tracking real satellites as compared with using simulated satellite signals.

The P-code receiver has been interfaced with a prototype antenna electronics unit and various jamming sources have been injected at the antenna to verify proper operation of both the antenna electronics and the GPS receiver. The jamming sources included CW, broadband, sweep, and pulsed CW, injected into the antenna at angles of 30 and 60 degrees with the GPS signal injected at an elevation of 90 degrees. During the testing, the jamming power was increased until the receiver lost lock. The jamming power and the signal power were then measured to determine the anti-jamming capability.

Test results with a single jammer, and with two jammers again prove that the anti-jam capability of the receiver and CRPA antenna meets predicted levels.

Dynamic Testing

An anti-jam dynamic test was performed on the breadboard P-code GPS receiver to verify its capability to resynchronize to the GPS signals after coming out of a jammed environment and having a position and velocity uncertainty due to IMU error. To perform this test, the velocity aiding normally received from the IMU was simulated. To fully simulate this aiding, the GPS software was programmed to measure pseudorange and delta pseudorange from an acquired satellite. With these measurements, the GPS processor can now predict satellite position and use these predictions to aid the receiver.

To simulate the uncertainty in the IMU due to a period of free inertial navigation, an error of 90 meters in position and 2 meters/second in velocity was incorporated into the aiding. To perform the test, a CW jamming source was combined with the GPS signals and the jamming signal turned up until the receiver lost lock. The jamming level was then reduced to the jamming threshold of the receiver and the time required to resynchronize to the GPS satellite measured. This took only 6 seconds. This type of test verifies the capability of the GPS receiver to be jammed off the air and be able to resynchronize to the GPS signal in a high jamming environment with the uncertainties predicted for the sensor package.

An anti-jam test was simulated to verify the increased anti-jam performance achievable from velocity aiding. During this test, a CW jamming source was inserted into the front end of the GPS receiver and the jamming level attenuated until the receiver was able to lock on a satellite signal. With the system running in a velocity-aided mode, the difference between the two levels of attenuation represents an improvement of 13 dB of AJ performance with velocity aiding. This is only 1 dB less than the figure often quoted as achieved with velocity-aided P-code receivers.

Acceleration tests were simulated to demonstrate and verify increased dynamic capability achievable with an integrated GPS/I system. To perform these tests, an HP9836 was used to simulate dynamics into the GPS/I system. The HP9836 was programmed with a flight profile which was used to drive the satellite simulator carrier frequency source from an HP3225A signal generator. The output of the signal generator was now a GPS signal whose carrier frequency was Doppler driven by the flight profile. The digital IMU sensor inputs were also driven by the HP9836 in place of the actual sensors. With both of these signals simulated, a flight profile of 0 g, followed by a high-g line-of-sight acceleration, and terminated with a 0 g profile was input to the GPS/I unit. In GPS-only mode, the receiver, because of large loop bandwidths, was able to maintain lock with accelerations up to 2.3 g. By using velocity aiding, the dynamic capability of this system was increased to 24 g.

Van Testing. After completion of lab testing, the GPS/I system was tested in a van along local interstate highways and in downtown Boston. The purpose of these tests was to evaluate performance with accelerations combined with angular motion. Figure 11 shows plots of one of these test runs. With GPS only, perturbations are found in the plot. These perturbations are caused by the effects of satellite blockages and acceleration changes on the GPS receiver. With the GPS/I system, these errors are removed. A blockage of approximately 30 percent was encountered during this test run.

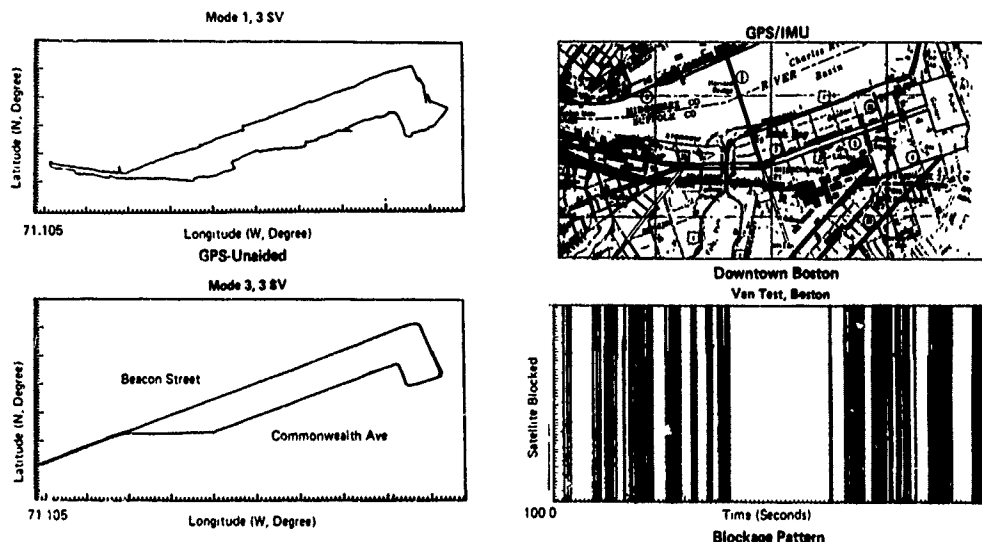


Figure 11. Satellite Reacquisition After Blockage

FLIGHT TESTING

The GPS/I unit was installed in a Navajo airplane and flight tested by comparing the position and velocity from this unit to that obtained by a single channel User Equipment receiver during flight, and to surveyed checkpoints on the ground. A base-line set of data from the GPS/I receiver taken prior to flight on the ground was shown to have an rms variation of only 2.1 ft and 0.3 fps. The latitude and longitude from the receiver is shown in Figure 12 over about one half hour of flight. The GPS antenna was connected after the plane was in flight and lock-on occurred within 2 minutes. Several tests were performed with the antenna disconnected for two minutes to force a period of free-inertial performance, then the antenna was reconnected. Recovery of the signals occurred within 10 seconds and the position error correction after all four SVs were recovered was about 10 ft. Lock was easily retained during the 2g turn. At return to the airport the position difference compared to the initial values was 20 feet.

IMU PERFORMANCE OPTIONS

After loss of all GPS satellites, navigation error growth is dominated by random gyro drift errors. Figure 13 shows free inertial position error growth for IMUs having typical error characteristics and three representative random gyro drifts. Selecting the proper gyro for integration into the GPS/I unit depends upon requirements for initial GPS acquisition, reacquisition after GPS signal loss, and performance at the end of a mission when very near a high-power jammer.

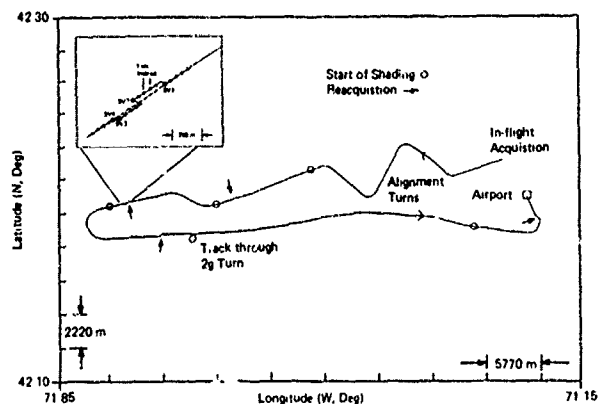


Figure 12. GPS/I Flight Tests

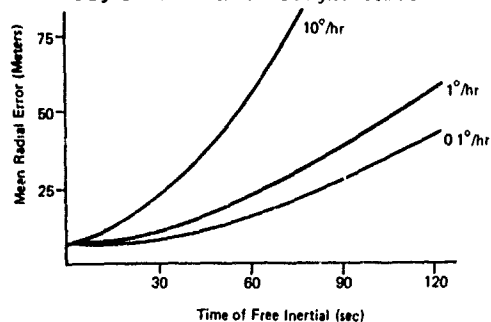


Figure 13. Free Inertial CEP After GPS Loss

ACKNOWLEDGEMENTS

The entire GPS receiver design team has contributed to the successful development of a high anti-jamming, miniature receiver with flexibility to be integrated with a variety of strapdown IMUs. This team found many innovative techniques to retain performance (accuracy, anti-jamming, fast TTFF, use with CRPA electronics, etc.) while reducing complexity, size, and cost.

REFERENCES

1. Maher, R., "A Comparison of Multichannel, Sequential and Multiplex GPS Receivers for Air Navigation," Texas Instruments, Inc., Lewisville, Texas 75067.
2. Divakaruni, S.P., "Ring Laser Gyro Inertial and GPS Integrated Navigation System for Commercial Aviation," IEEE PLANS, 1986, p 73.
3. Teasley, S.P., "Flight Test results of an Integrated GPS and Strapdown Inertial System," IEEE PLANS, 1986, p 199.
4. Nielson, J.T., "GPS Aided Inertial Navigation," NAECON, IEEE 1986, p 20.
5. Turin, G. L., "An Introduction to Digital Matched Filters," Proc. IEEE, Vol. 64, No. 7, July 1976, pp 1072-1112.
6. Amoroso, F., "Adaptive A/D Converter to Suppress CW Interference in Spread-Spectrum Communications," IEEE Trans. Comm., Vol. COM-31, No. 10, Oct. 1983, pp 1117-1123.
7. Natali, F. D., "All Digital Coherent Demodulator Techniques," Proc. International Telemetry Conference, Vol. VIII, Oct. 1972, pp 89-108.
8. Murphy, J. W. and M. D. Yakos, "Collins Avionics NAVSTAR GPS Advanced Digital Receiver," Conf. Record of National Aerospace Meeting of the Institute of Navigation, Arlington, Virginia, 23-25 March 1983, pp 27-36.
9. Spilker, J. J., "Digital Communications by Satellite," Prentice-Hall, Englewood Cliffs, New Jersey, 1977.

APPLICATION OF MULTIFUNCTION INERTIAL REFERENCE SYSTEMS TO FIGHTER AIRCRAFT

by

Carlos A. Bedoya, McDonnell Douglas Corp.
John M. Perdsock, Air Force
Wright Aeronautical Laboratories
Wright Patterson AFB
OH 45433-6553
United States

Introduction

As requirements for Flight Control, Fire Control, Propulsion Control and Navigation Systems are developed for future fighter aircraft, reliability, maintainability, availability, redundancy, and survivability become key issues. These systems require dependable and accurate sources of inertial measurement data. The Multifunction Flight Control Reference System (MFCRS) was developed to demonstrate the use of a minimum number of inertial sensors in a survivable configuration to provide inertial data for flight control, navigation, weapon delivery, cockpit displays, and sensor stabilization.

The MFCRS Program used two extensively modified Ring Laser Gyro (RLG) navigation units, developed originally for the AV-8B aircraft by Honeywell Inc., to perform the flight control reference and navigation functions on board an F-15 fighter aircraft. These two motion reference units (MRU's) were separated by nine feet in the aircraft to provide survivability and skewed to provide redundant inertial information. This chapter will give an overview of the various stages of development that have been completed on this program, the lessons learned to date, and what is planned for the future. The following paragraphs give a summary of the items to be covered.

Evaluation of MFCRS performance and suitability for installation in the F-15 required an extensive laboratory test and integration effort at McDonnell Aircraft Co. (MCAIR) in St. Louis, MO between May and December 1983. The evaluation was primarily performed in the MCAIR Navigation Systems Laboratory, with the system being interconnected to the F-15 flight control system in the MCAIR Flight Control Laboratory during the integration portion of the testing. The laboratory testing was undertaken to evaluate performance at the system level and evaluate flight control outputs, redundancy management, electronic MRU to MRU alignment, reaction time, navigation performance, performance under vibration, temperature, EMI environments, and operation when integrated with the F-15 Flight Control System. The main objective of this testing was to determine the suitability of this system for installation and flight test in the F-15.

Following the MFCRS laboratory evaluation a ground structural mode interaction test and a two phase flight test program was performed. The objective of the first flight test phase was to verify MFCRS air worthiness, to compare and evaluate MFCRS flying qualities with the flying qualities of the basic F-15, to verify proper MFCRS redundancy management operation, and to verify that the MFCRS sensors were of navigation quality. The phase one flight testing conducted during February 1984 evaluated the MFCRS redundancy management operation, revealed a low damping problem in the MFCRS Control System response at medium to high dynamic pressure flight conditions, and verified the navigation accuracy of the normal and skewed sensors. When the MFCRS low damping problem appeared it became necessary to account for the differences between expected and actual system performance. The low damping analysis identified the problems to be time delays present in the system. The second phase of the flight test program was conducted in July 1984 with time delays reduced within the capability of the system. The details and results of both phases of the flight test program, together with the causes of the low damping problem will be discussed in the chapter.

Subsequent to the successful completion of the phase two flight test evaluation, changes to the MFCRS hardware and software structure, which would improve system performance and expand the MFCRS flight envelope, were identified. These changes were designed and evaluated as part of the "Enhanced MFCRS" (EMFCRS) study which began in the fall of 1984. The EMFCRS Program involved the development, implementation, and laboratory evaluation of the necessary hardware and software changes to expand the MFCRS flight envelope. Following the laboratory evaluation additional flight testing was planned for the fall of 1986 to verify that the EMFCRS configuration would result in level 1 handling qualities in both supersonic and subsonic flight as well as in tracking of target aircraft. Unfortunately, during ground testing prior to flight, a 22 hz structural mode interaction was found in the control system pitch channel. This mode was unexpected as it had not been observed previously and was not in the available structural model. The system changes developed during the EMFCRS studies, the laboratory test results, and the aircraft testing will be discussed, as well as the follow on development of multifunction systems now in process under the name of Ada Based Integrated Controls System (ABICS) Phase III.

Background

Current operational fighter aircraft contain several sets of inertial sensing equipment including Inertial Navigation Systems (INS), Attitude and Heading Reference Systems (AHRS), Flight Control Gyros and Accelerometers, and Fire Control Lead Computing Gyros, each of which is dedicated to, and optimized for, a specific application. Possible improvements in cost, weight, size, reliability, and maintainability as a result of combining some of these sensor sets were studied under the Multifunction Inertial Reference Assembly (MIRA) program which was jointly sponsored by Air Force Wright Aeronautical Laboratories Flight Dynamics Lab (AFWAL/FI) and Avionics Lab (AFWAL/AA).

*Ada is a registered trademark of the U.S. Government (Ada Joint Program Office).

The AFVAL Flight Dynamics Laboratory subsequently sponsored the Multifunction Flight Control Reference System (MFCRS) program to develop and flight demonstrate the flight control aspects of MIRA using strapped down inertial quality sensors as a systems reference for an advanced flight control system operating in a dynamic fighter aircraft.

The key technical areas investigated during the MFCRS program were:

- o The suitability of navigation quality ring laser gyros and accelerometers, in a strapped down configuration, for use as flight control reference sensors.
- o Control law compensation for clustered sensors (gyros and accelerometers) and location effects.
- o Redundancy management techniques associated with skewed and dispersed sensor clusters.

General System Description

The MFCRS equipment consisted of two modified AV-8B H421 Laser Inertial Navigation Systems (LINS), and one Test Management Panel (TMP). The modified AV-8B LINS unit is called a Motion Reference Unit (MRU) in the MFCRS program. This equipment provides dual navigation outputs as well as navigation grade sensor outputs for use in the F-15 flight control system. Figure 1 depicts the MFCRS equipment and its signal interfaces. The actual hardware is shown in Figure 2.

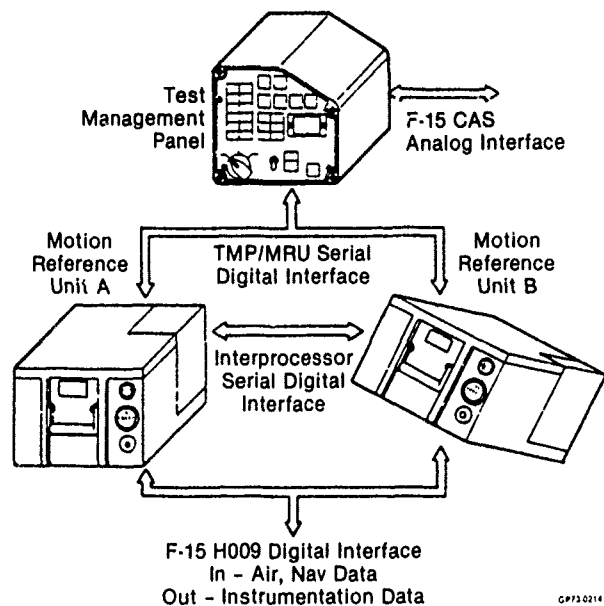


Figure 1. MFCRS Equipment

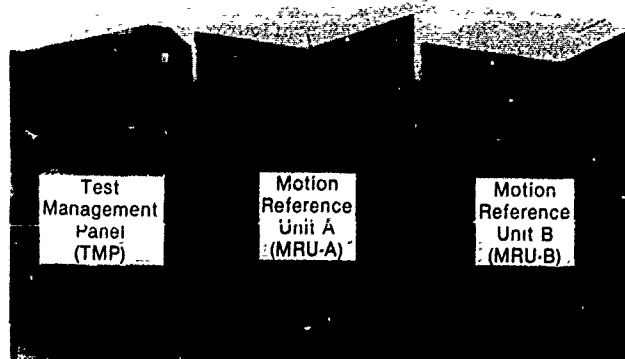


Figure 2. MFCRS Units

The MFCRS equipment was physically located in the F-15 aircraft as depicted in Figure 3. MRU-A was installed in the forward nose section, MRU-B was installed in the No. 5 equipment bay and the TMP was installed in the upper right corner of the main instrument panel.

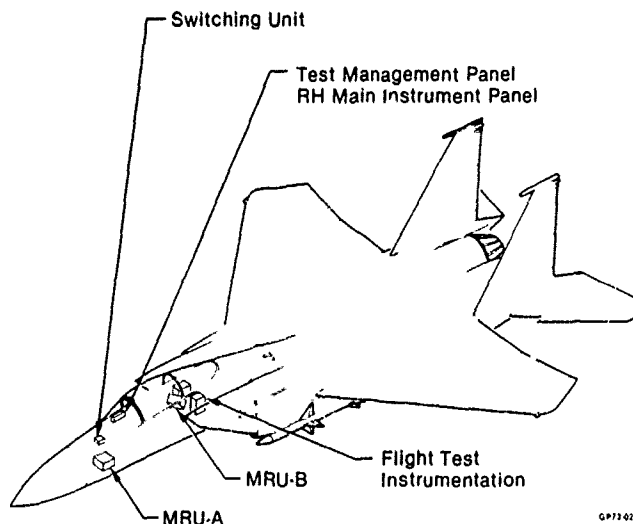


Figure 3. MFCRS Equipment Location

For survivability the sensor packages of this type of system would be separated in the aircraft. Previous studies have shown that this separation should be about 30 inches. Due to limited locations in the F-15 test airplane large enough to accommodate the MRUs, it was necessary to separate them by 9 feet. The forward sensor package is aligned with the aircraft body axes. The aft sensor package is skewed 60° about the cone axis of the MRU-A sensor. Figure 4 shows the equipment installed in the test aircraft.

The size, weight, and power of the MFCRS equipment is shown in TABLE 1.

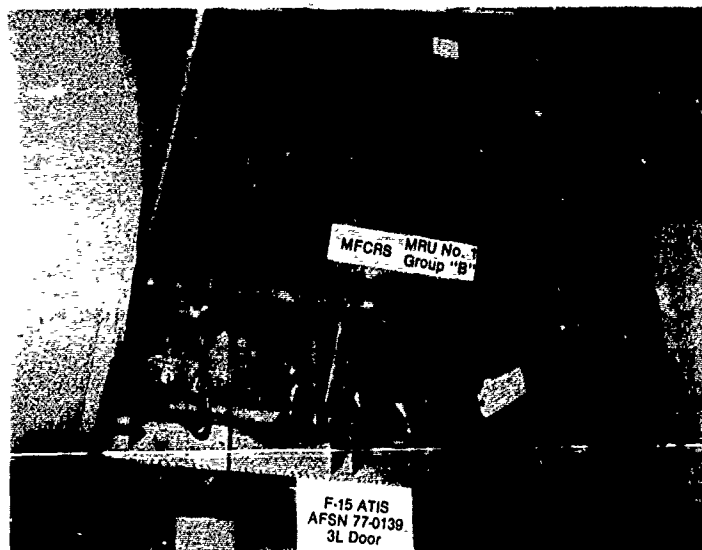


Figure 4A. MFCRS and ATIS Installation in F-15

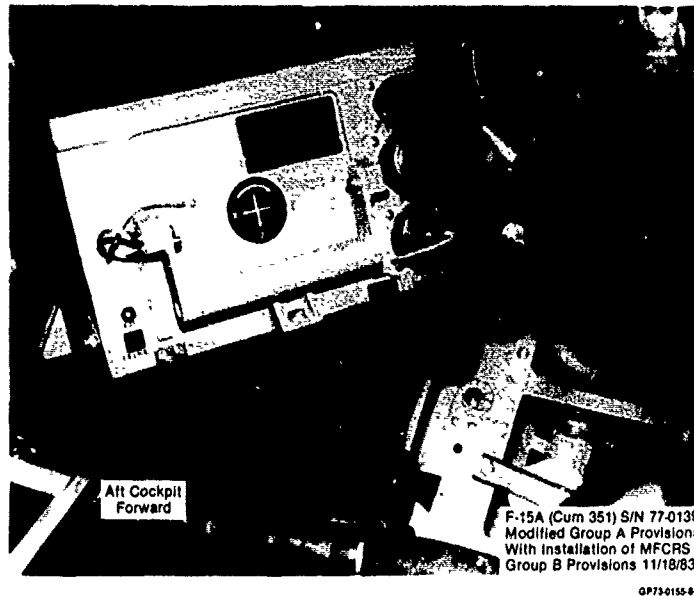


Figure 4B. MRU-B Skewed Installation in the F-15

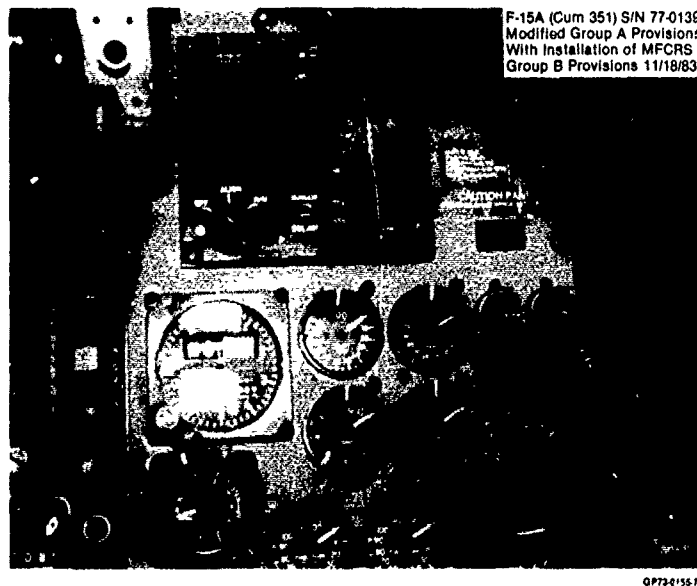


Figure 4C. MFCRS Test Management Panel Installation in the F-15

TABLE 1. MFCRS EQUIPMENT SIZE, WEIGHT AND POWER

Unit	Size (in.)	Weight (lb)	Power (watts)
MRU-A, -B	7.6 H x 14 L x 11 W	49	120
TMP	5.3 H x 15.5 L x 6 W	16	32

The MRU contains modifications to: 1) permit digital data to be interchanged with both the other MRU and the TMP, 2) synchronize the digital processing in both MRUs, and 3) permit digital data to be interchanged with the F-15 mission computer via the H009 data bus.

The TMP was a new design specifically for MFCRS. It contains electronics to: 1) permit exchange of digital data with both MRUs, 2) provide an interface with the pilot for mode control, in-flight fault simulation, and status annunciation, 3) provide backup battery to power the MFCRS during momentary aircraft power interrupt, and 4) convert aircraft flight control signals from digital to analog form. The front panel of the TMP is shown in Figure 5.

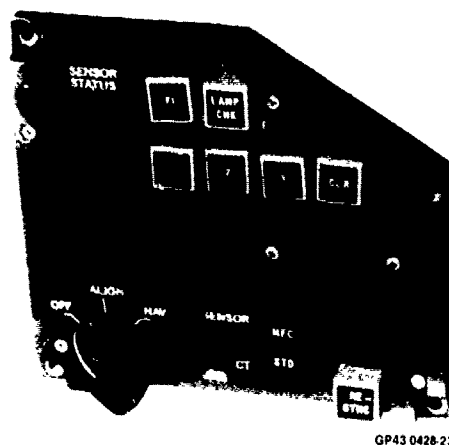


Figure 5. Test Management Panel (TMP)

The TMP flight control outputs consist of 400 Hz AC roll, yaw, and pitch rate outputs and DC level lateral and normal acceleration. There is also a 26V, 400 Hz Rate Gyro connector interlock signal which is serially disconnected when a fault condition occurs to switch over to the standard sensors.

A functional block diagram of the F-15 flight control system using MFCRS sensors is shown in Figure 6. The orientation of the MFCRS sensors with respect to the aircraft axes is illustrated in Figure 7.

MFCRS Control System

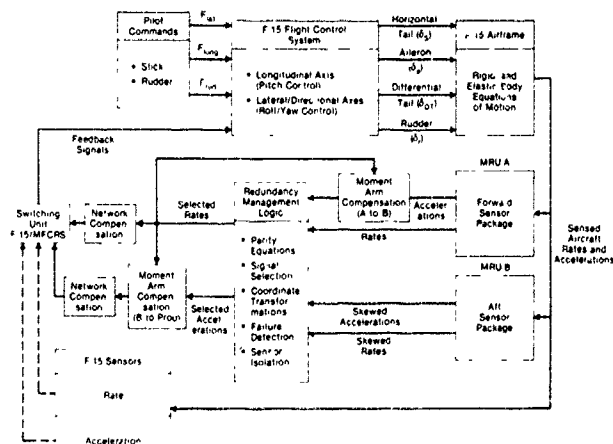
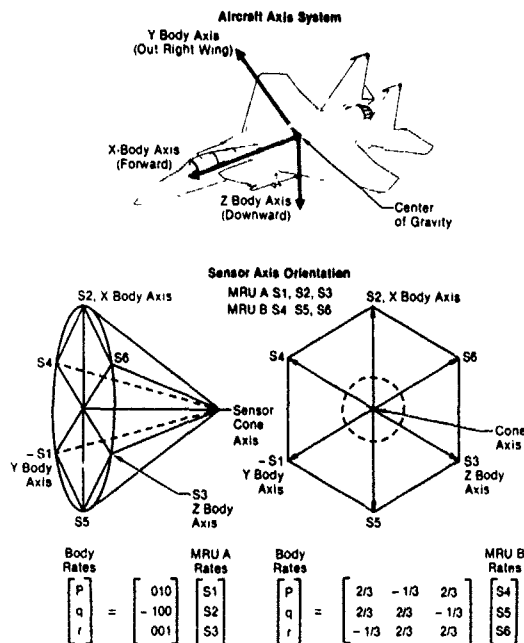


Figure 6. MFCRS Control System

The F-15 flight control system is an analog fixed gain dual channel control augmentation system (CAS) with a mechanical control system operating in parallel. The elements of this control system were not modified in the MFCRS studies; the structural mode and moment arm compensation required for MFCRS are added to the sensor feedback path.



GP720214 15

Figure 7. MFCRS Sensor Axis System Orientation

The MFCRS control system was designed to maintain basic F-15 stability and performance characteristics when integrated with an unmodified F-15 control augmentation system (CAS). Integration with the existing CAS made it necessary to design the MFCRS structural mode compensation around the compensation in the CAS, and required that the compensation needed to offset MFCRS system lags (sensor and computational) be located in the control system feedback paths. This approach proved to be one of the shortcomings that hampered control system design.

Several factors in addition to the design of the existing CAS, the military specifications applicable to the F-15/MFCRS and pilot evaluation during the manned simulation were considered in the design of the MFCRS flight control system. These factors are:

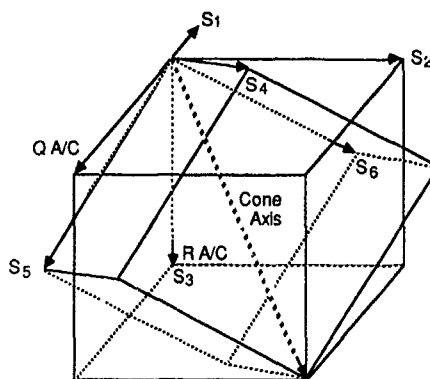
- 1) Unique sensor characteristics (such as mechanically dithered ring laser gyros)
- 2) System lags resulting from digital computations, A/D-D/A conversions, and network compensation
- 3) Variations in aircraft structural mode frequencies with flight condition and stores configuration
- 4) Sensor location effects (moment arm corrections)
- 5) Sensor assembly separation

The Redundancy Management (RM) System developed for MFCRS provides fail-op, fail-op, fail-safe capability for the inertial sensors and dual redundancy for the remainder of the system including the F-15 Control Augmentation System (CAS). The MFCRS redundancy management system operates to provide the best three gyro and accelerometer outputs to the F-15 flight control system. For MFCRS using six sensors, three at a time, the redundancy management system selects the best sensor triad from among the 20 possible triad combinations. The F-15 CAS contains cross-channel monitoring circuitry which will cause reversion to the mechanical back-up system if differences between the two input channels exceed acceptable levels. Dual channel isolation and monitoring consistent with prudent design practices were incorporated in the MFCRS design.

The forward MRU (MRU-A) was installed in the nose section of the F-15 and aligned with the aircraft pitch, roll, and yaw axes. The aft MRU (MRU-B) was installed approximately nine feet aft of MRU-A in a skewed position such that its sensor triad was rotated 60° about the cone axis of the MRU-A sensor triad. The resulting sensor geometry is shown in Figure 8 and is defined by:

$$\begin{matrix} Q \\ P \\ R \end{matrix} = \begin{bmatrix} -1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \times \begin{bmatrix} S_1 \\ S_2 \\ S_3 \end{bmatrix} = \begin{bmatrix} +2/3 & +2/3 & -1/3 \\ +2/3 & -1/3 & +2/3 \\ -1/3 & +2/3 & +2/3 \end{bmatrix} \times \begin{bmatrix} S_4 \\ S_5 \\ S_6 \end{bmatrix}$$

where: Q = aircraft pitch axis
P = aircraft roll axis
R = aircraft yaw axis
 S_1, S_2, S_3 = MRU-A sensor axes
 S_4, S_5, S_6 = MRU-B sensor axes



GP73-0524 3-D

Figure 8. Sensor Axis Orientation

The MFCRS installation is considered characteristic of a "worst case" installation from the standpoint of sensor location and separation. The development of effective sensor compensation and redundancy management algorithms for such an installation was essential in order to demonstrate the suitability of strapped-down inertial sensors for use as flight control feedback sensors.

There are a large number of redundancy management approaches discussed in current literature. The approaches include use of parity equations, observers, analytic redundancy, Kalman filters, and generalized likelihood tests. Unfortunately, required numerical accuracies and execution times preclude the use of some of these algorithms in a real-time system implemented in a small on-board computer, especially when other functions must be performed. The Multifunction Flight Control Reference System (MFCRS) contains four processors with each MRU having a complete set of flight control, navigation, and redundancy management algorithms in one processor and a second processor performing the filtering of the gyro dither noise. The candidate offering the simplest implementation, fastest execution and smallest memory requirement is redundancy management with parity equations in conjunction with a look up table.

The number of parity equations available for use by the redundancy management logic is equal to the number of combinations of N sensors when taken four at a time. This is expressed by:

$$(N, R) = \frac{N!}{R!(N-R)!}$$

where: R = number of sensors used in each parity equation (4)
N = number of sensors available

The complete MFCRS parity equation set is:

$$P_n = \begin{bmatrix} -2 & 2 & -1 & -3 & 0 & 0 \\ -2 & -1 & 2 & 0 & -3 & 0 \\ 1 & 2 & 2 & 0 & 0 & -3 \\ -2A & A & 0 & -2A & -A & 0 \\ A & -2A & 0 & 2A & 0 & A \\ B & B & 0 & 0 & B & -B \\ -2A & 0 & A & -A & -2A & 0 \\ B & 0 & B & B & 0 & -B \\ A & 0 & -2A & 0 & 2A & 0 \\ -3 & 0 & 0 & -2 & -2 & 1 \\ 0 & -B & B & B & -B & 0 \\ 0 & -2A & -A & A & 0 & 2A \\ 0 & -A & -2A & 0 & A & 2A \\ 0 & -3 & 0 & 2 & -1 & 2 \\ 0 & 0 & -3 & -1 & 2 & 2 \end{bmatrix} \times \begin{bmatrix} S_1 \\ S_2 \\ S_3 \\ S_4 \\ S_5 \\ S_6 \end{bmatrix}$$

where: $A = \frac{3\sqrt{5}}{5}$ and

$$B = \frac{3\sqrt{2}}{2}$$

The sensor compensation techniques developed to reduce sensor errors caused by sensor misalignments, bias, static bending, dither noise, and moment arm effects are discussed below:

o Sensor Misalignments

The precision to which the MRUs are installed relative to the aircraft body axes and to each other impacts both output accuracies and the performance of the MFCRS redundancy management logic.

Differences between theoretical and actual sensor orientations will result in errors in the MFCRS flight control outputs. The misalignment errors that show up in the MFCRS redundancy management logic after the sensor data has been compared are caused by relative misalignments between the MRUs. Because redundancy management decisions are based on the magnitude of sensor differences, errors due to relative sensor misalignments will reduce the sensitivity of the logic by requiring higher decision thresholds during dynamic flight. Three steps are taken for MFCRS to insure that sensor misalignment errors are kept acceptably small.

First, the MFCRS MRUs are modified AV-8B inertial navigation units. Very tight tolerances are maintained during the AV-8B manufacturing process for sensor-to-sensor alignments within the orthogonal sensor cluster in each MRU and for the alignment of the sensor cluster to the MRU chassis.

The second step involves the accurate installation of the MRUs in the test aircraft. The installation location and normal orientation of the forward MRU (MRU-A) allows the use of mechanical boresighting techniques. The location and skewed orientation of the aft MRU (MRU-B) does not allow the use of conventional boresighting techniques to obtain a precision installation.

The limited MRU to MRU sensor alignment accuracies achievable in the MFCRS installation by physical alignment techniques did not support the development of an effective and efficient redundancy management approach. The need to improve these accuracies led to the development of a significant new technique for MFCRS, electronic alignment.

The third step, electronic alignment of the MFCRS installation, is accomplished by using the navigation capability of each MRU, after the MRUs have been installed in the test aircraft. The electronic alignment process will be performed only once during the MFCRS program with the results being stored in nonvolatile memory in each MRU.

To perform the electronic alignment the system is turned on in align and allowed to thermally stabilize. Multiple alignments are then performed. The aircraft is then rotated 180 degrees in heading and additional alignments are performed. From these alignments a resultant transformation (T) matrix is obtained which accurately specifies the orientation of MRU-A and MRU-B to each other.

o Parity Equation Bias Removal

Sensor bias errors are linearly combined in the MFCRS parity equations and appear as bias errors in the parity equation outputs. A parity equation bias removal routine was developed to eliminate steady state sensor errors from the MFCRS parity equations. The values of the parity equations are determined under static conditions during system initialization and used as parity equation bias terms. In addition to these initial bias terms, correction terms are also determined during nonmaneuvering flight conditions to account for any changes in bias as a function of time. The bias removal terms are added to the parity equations each time they are computed, effectively cancelling any steady state bias.

o Static Bending Misalignments

Sensor misalignment errors that are caused by aircraft static bending become significant only during maneuvering flight - as the aircraft structure flexes. Since pilot sensitivity to sensor switching transients is reduced due to increased aircraft structural background noise during maneuvering flight, it was possible to effectively mask structural bending misalignment errors from the redundancy management logic by the use of scheduled fault detection and sensor selection thresholds (trip levels).

o Dither Noise

The Ring Laser Gyros used in the MFCRS are dithered to prevent "lock in" at low angular inputs. This dither signal is aliased into the 1hz to 25hz range and

appears on both the accelerometer and gyro outputs as noise. To suppress this noise a -60db digital notch prefilter at the dither frequencies was placed in the gyro path and a third order analog lag prefilter was placed in the accelerometer path. A 0.1 second lag filter was also used to meet F-15 control signal output noise specifications.

o Moment Arm Effects

Since the two MRUs were separated by nine feet the sensed acceleration at MRU-A is not the same as the acceleration at MRU-B. In order to compare the outputs for redundancy management the MRU-A and MRU-B accelerations are compensated to the production F-15 flight control sensor location for redundancy management and to maintain F-15 handling qualities.

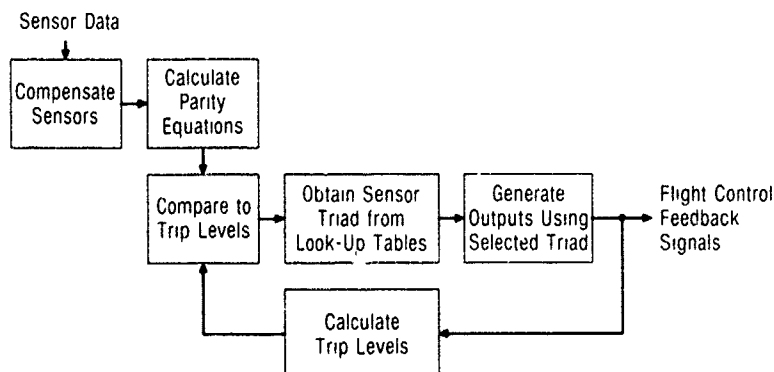
o Redundancy Management

The method used in the MFCRS to select the sensors to be used to provide the flight control outputs was a table lookup. The table look-up logic computes sets of parity equations (selected as a function of the sensor failure status) and compares the value of the parity equations to two trip levels, one for sensor selection and one for fault detection and isolation. Precomputed redundancy management decisions are obtained from stored tables using table pointers generated by the parity equation/trip level comparisons.

Salient features of this approach include:

- o Minimum processor utilization (RM decisions are computed off-line and stored in look-up tables).
- o Ability to deal with dual, simultaneous features.
- o Noise immunity (via use of trip levels).
- o Flexibility (look-up tables and trip levels are easily adjusted).
- o Two level operation - acceptable sensitivity without false alarms.
- o Decisions based on status of all parity equations computed.

Real-time operation of the redundancy management logic developed for MFCRS is shown in Figure 9.



GP43-0428-19

Figure 9. MFCRS Redundancy Management Logic
Simplified Block Diagram

Sensor selection decisions are made every sample period (0.02 seconds), with all valid sensors being considered. Fault detection and isolation decisions are based on information obtained over several sample periods. Once a sensor failure has been isolated, the faulty sensor is removed from further consideration by the redundancy management logic.

In addition to the flight control functions described the MFCRS also mechanized full navigation capability in both MRUs. The normal navigation outputs of position, velocity, and attitude are supplied from each MRU. These signals were not displayed in the aircraft but were recorded on instrumentation so that navigation performance could be determined for each flight. The navigation output accuracy goals are shown in TABLE 2.

TABLE 2. MFCRS NAVIGATION OUTPUT ACCURACY GOALS

Parameter	MRU-A	MRU-B	Range	Units
North and East Velocity	12 00	18 00	$\pm 3,200$	ft/sec
North and East Acceleration	3 00	3 00	± 322	ft/sec ²
Inertial Altitude	50 00	50 00	- 1,000 + 55,000	ft
Vertical Velocity	6 00	9 00	$\pm 1,500$	ft/sec
Vertical Acceleration	3 00	3 00	± 322	ft/sec ²
True Heading	0 22	0 32	± 180	deg
Pitch and Roll Attitude	0 15	0 20	$\pm 90 / \pm 180$	deg
Present Position	1 00	3 00	Note 1	NM/hr (CEP)

Note 1 Latitude $\pm 90^\circ$, longitude $\pm 180^\circ$

00750214-011

Laboratory Evaluation

In order to evaluate MFCRS performance and suitability for installation in the F-15 an extensive laboratory test and integration effort was undertaken at McDonnell Aircraft Co. (MCAIR) in St. Louis, MO. The evaluation was primarily performed in the MCAIR Navigation Systems Laboratory, with the system being interconnected to the F-15 flight control system in the MCAIR Flight Control Laboratory during the integration portion of the testing.

The objective of the testing was to evaluate performance at the system level and evaluate the following:

- o Flight control outputs
 - Steady state rate outputs
 - Sinusoidal rate outputs
 - Acceleration outputs
- o Redundancy management
 - Static conditions
 - Dynamic conditions
- o Built in test
- o Electronic MRU to MRU alignment
- o TMP operation
- o Operation with scorsby motion applied
- o Operation of Sensor Switching Unit
- o Performance under vibration, temperature and EMI environments
- o Power Consumption
- o Operation when integrated with the F-15 Flight Control System
- o System interface with the H009 bus
- o Reaction Time
- o Navigation performance

System Testing

An extensive amount of system and integration testing was performed on the MFCRS in the Navigation Systems Laboratory of MCAIR in St. Louis, MO between May and December 1983. Figure 10 shows the MFCRS MRUs mounted on the two-axis GOERZ table in the MCAIR avionics laboratory. The main objective of this testing was to determine the suitability of this system for subsequent installation and flight test in the F-15. The following tests were performed:

- o ELECTRONIC ALIGNMENT
- o INTEGRATION TESTS
- o STATIC SYSTEM NOISE TESTS

During the first flight control integration test with actuators, it became apparent that the noise specification tolerances, 20mv for accels and 65mv total for rates, were not correct. Even though the MFCRS output noise levels were less than or equal to the standard flight control sensors, MFCRS outputs caused noticeable movements on the actuators due to the low frequency spectrum in these outputs. We found that noise levels as low as 2 milli-volts in the actuator bandwidth could cause actuator motion.

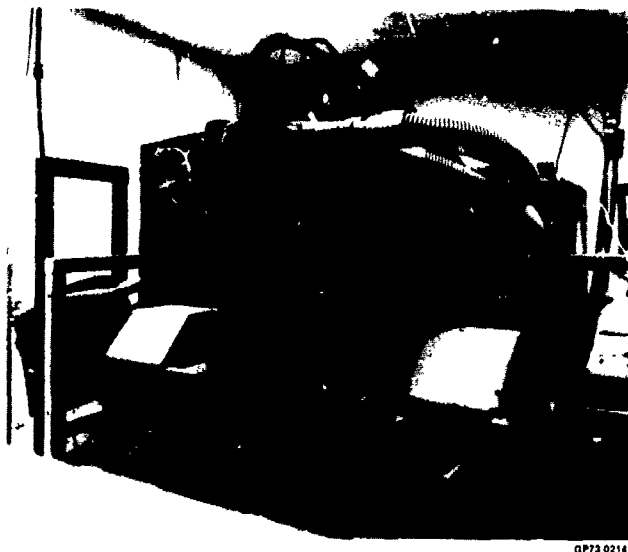


Figure 10. MFCRS Mounted on Two-Axis Goerz Table in Yaw Configuration

Several ideas were discussed and investigated on how to reduce the noise levels. It was decided that the noise levels would have to be reduced low enough to cause less than 0.2° actuator movement. A combination of rescaling the accelerometer outputs and modifying the redundancy management look up tables was used to reduce actuator motion to within acceptable limits.

o **REDUNDANCY MANAGEMENT TESTING**

During these tests the Flight Control Actuators were monitored on recorders as well as visually.

The results of these tests did show transients and actuator motion when changing sensors but all of these transients were low enough as not to affect the flight control system operation.

o **SINUSOIDAL (GAIN AND PHASE) RATE TESTS**

o **SCORSBY TESTING**

Laboratory Test Conclusions

- o One of the lessons learned is that for flight control, the sensor noise tolerance should be specified as a function of frequency and not just as a peak to peak amplitude.
- o Unique laboratory testing techniques were developed for the MFCRS testing such as the MFCRS fixture to hold both measurement reference units in relative orientation to each other. This fixture proved to be invaluable during rate testing, system calibration, attitude testing, redundancy management tests and sinusoidal response tests.
- o The concept and procedure for electronic alignment, which uses the navigation alignment capability of MRUs to determine their orientation relative to each other was proven to be more accurate than anticipated.
- o The navigation accuracy of both the conventionally mounted MRU and skewed MRU exceeded the current F-15 requirement.
- o Integrated systems such as MFCRS require that systems such as flight control, navigation, and aircraft central computer be integrated together in order to effectively evaluate overall system performance.
- o Sinusoidal rate testing is a very effective method for quickly showing up timing problems that may exist when using combinations of sensors from different boxes. This testing is essential in all axes with systems such as MFCRS.
- o Sinusoidal rate testing is also a necessary test in order to verify the gain and phase margin specifications of the flight control outputs.

- o The extensive integration testing with MFCRS and the flight control system with the aircraft rudder and horizontal stabilator actuators revealed problems which would otherwise have been very difficult to see. In addition this overall integration testing proved that in an open loop sense the MFCRS could provide the rate and acceleration signals needed by the flight control system in the F-15 aircraft.

Flight Testing

Following the laboratory evaluation the MFCRS was installed in an F-15 aircraft and a structural mode interaction test and flight test program were undertaken. This testing was performed between December 1983 and July 1984. The actual flight testing was conducted in two phases with Phase 1 in February 1984 and Phase 2 in July 1984. Between February and July the results of the Phase 1 flight test were evaluated, system changes made, and additional laboratory testing performed to prepare for Phase 2 of the flight test program. These phases will be individually discussed in the following sections.

PHASE 1 FLIGHT TEST

The specific objectives and success criteria for the Phase 1 flight test program were:

Flight Test Objectives

- a) To verify MFCRS Airworthiness.
- b) To compare and evaluate aircraft flying qualities using MFCRS feedback sensors with the flying qualities of the basic F-15.
- c) To evaluate the MFCRS at representative flight conditions for susceptibility to false alarms.
- d) To verify proper MFCRS redundancy management operation.
- e) To verify that the MFCRS sensors are navigation quality.

Flight Test Success Criteria

- a) MFCRS operates safely - no unacceptable system transients or aircraft anomalies system transients or aircraft anomalies occur.
- b) MFCRS operation does not degrade aircraft stability and control. Changes in aircraft handling characteristics are not apparent to the pilot.
- c) MFCRS is not susceptible to false alarms as a result of variations in flight condition or sensor configurations.
- d) Faulty sensors are identified and removed from the system. Sensor switching/reconfiguration transients are acceptably small.
- e) Terminal position and velocity accuracies are equal to or better than the basic F-15 INS specification for MRU-A and MRU-B.

Following a functional check flight in St. Louis the test aircraft was ferried to Edwards AFB to begin the MFCRS AFFTC/MCAIR flight test. The first airworthiness flight at Edwards was flown with MFCRS generating flight control and navigation outputs for real time and post flight analysis, but with standard sensors providing flight control information to the CAS. The next two airworthiness flights were flown with successful MFCRS engagement. During the next two flights system damping was found to be unacceptable for closed-loop tasks at medium to high dynamic pressure flight conditions, and further flying quality evaluations were suspended. A redundancy management flight and a structural mode flight to gather data for troubleshooting of the damping problem were flown before the MFCRS components were removed from the aircraft on 1 March 1984.

MFCRS AIRWORTHINESS - The primary objective of the initial MFCRS flights was to verify MFCRS airworthiness. MFCRS sensor engagements/disengagements, small amplitude maneuvers (stick raps and rudder kicks), rolling pull-ups, and frequency sweeps were performed to test for MFCRS induced aircraft transients, structural oscillations, or other abnormal aircraft responses. The MFCRS was evaluated through real time monitoring of safety-of-flight measurands, as well as post flight analysis of on-board data and pilot comments.

Unacceptable damping was noted following small amplitude maneuvers and frequency sweeps in all three control axes with MFCRS sensors engaged at the highest q conditions tested (0.9 Mach/15K Ft, 0.9 Mach/7.5K Ft). Noticeably degraded damping was observed at more moderate q flight conditions (0.6 Mach/15K Ft). Also, during Flight 5 a pass through turbulent air during 1g flight at 0.85 Mach/7.5K Ft with MFCRS sensors engaged resulted in increasing aircraft oscillations to the point where the pilot disengaged MFCRS with the control stick disengage switch. The aircraft stabilized upon MFCRS disengagement and automatic CAS reengagement.

No objectionable aircraft transients or other unusual behavior were observed at all test points during MFCRS 1g sensor engagements, 1g to 3g sensor disengagements, and rolling pull-up maneuvers. Unexpected triad switching was noted in very benign flight conditions during Flight 4, but was corrected on subsequent flights by raising sensor select trip levels through an EPROM change.

MFCRS FLYING QUALITIES - Aircraft flying qualities with MFCRS sensors engaged were evaluated in conjunction with airworthiness testing. Large Amplitude Maneuvers (LAM's) (barrel rolls, loops, cloverleaves) were performed in addition to airworthiness maneuvers to qualitatively evaluate aircraft stability/control and handling characteristics. MFCRS triads 456, 123, 134 and 246 were tested.

As mentioned, aircraft damping was unacceptable with MFCRS sensors engaged at the highest q flight conditions tested. However, aircraft flying qualities with MFCRS sensors were satisfactory at relatively low Q conditions (0.6 Mach/30K Ft) and during large amplitude maneuvers with slow, smooth control inputs. No observable difference in aircraft response was noted between any of the four triads tested.

FALSE ALARMS - MFCRS susceptibility to false alarms was monitored during all MFCRS flights. False alarms are indications on the TMP that a MFCRS sensor had malfunctioned or failed, when in fact it has not.

A total of five false alarms were observed on the TMP during the first seven flights. Each was cleared with the TMP CLR (Clear) pushbutton. The number of strikes required for fault detection were increased to 5 for the first two failures and 3 for the third failure. There were no more false alarms after this change.

REDUNDANCY MANAGEMENT - Flight 6 was dedicated to verification of proper MFCRS redundancy management operation. Redundancy management refers to the ability of MFCRS to assimilate redundant flight control information, detect sensor malfunctions or failures, and reconfigure to the best available triad. Redundancy management testing during flight 6 consisted of inputting three sequences of five simulated sensor faults of four types (hardover, null, bias, and scale factor) to MFCRS. Faults were inserted with MFCRS engaged to the flight control system during straight/level, rolling pull-up, and single-axis maneuvers at 0.8 Mach/30K Ft. Frequency sweeps about all three axes were also performed after fault insertion. MFCRS was then monitored for proper fault detection, triad reconfiguration, CAS disengages, and aircraft transients.

All simulated sensor faults were correctly detected by the system, and there were no false alarm faults detected using the revised fault detection criteria. Mild aircraft transients were noted after insertion of faults A2 (acceleration hardover) and G5 (gyro bias). CAS disengages were also noted in conjunction with automatic MFCRS disengagement after insertion of the fifth fault.

SENSOR RECONFIGURATION TRANSIENTS - When the MFCRS redundancy management logic changes the triads for generating the flight control outputs and the system switches, a step change will occur in the MFCRS flight control outputs which will be equal to the difference between the output solutions using the old and new triads.

Normally occurring sensor differences during static flight were expected to be small and reconfiguration should not result in detectable transients. However, sensor switching transients are likely to be larger during maneuvering flight due to aircraft dynamic bending effects.

As the aircraft bends during maneuvering flight, the sensor cluster in one MRU is tipped relative to the cluster in the other. This tipping causes the relative alignment of the sensors to change, resulting in difference in the MFCRS outputs (pitch rate, roll rate, yaw rate, normal acceleration and lateral acceleration) as a function of which sensor triads are selected for use in generating the flight control outputs.

The magnitude of sensor switching transients caused by static bending is a function of: (1) the maneuver, (2) the triad selected prior to sensor switching, and (3) the triad selected after sensor switching. Figure 13 summarizes the results of a simulation study conducted to determine the relationship between static bending and sensor switching transients. The misalignment signals (δ 's) developed during the simulated maneuver are shown in each axis for all combinations of sensors. The transitions between sensor triads that will be determined by finding the two sensor triads in whose δ 's differ the least. Likewise, the worst case transients will occur when switching between sensor triads whose δ 's differ the most. Sensor switching transients were evaluated in-flight on an axis-by-axis basis for a variety of pitch, roll and yaw maneuvers.

NAVIGATION - MFCRS navigation data was recorded from the MCI following all flights (except ferry), and also following a pilot proficiency flight. MFCRS retains its original function of generating navigation information, but standard F15 INS (ASN109) navigation outputs are used by aircraft systems for navigation purposes. The sole navigation objective of the MFCRS flight test program was to verify navigation quality of MFCRS sensors by acquiring MRU-A and MRU-B terminal position and velocity accuracies equal to or better than basic F-15 INS specification.

Two MFCRS alignment options were available: an automatic 4 min 30 sec alignment or an extended, manually controlled alignment can be selected. The automatic 4 min 30 sec alignment was used on flights 6, 7, shakedown, and pilot proficiency. A manually controlled alignment of approximately 9 min 30 sec was used on all other flights.

Valid navigation data was obtained on all flights, with possibly two exceptions. The nav data obtained from MRU-A following flight 6 contained large latitude and velocity errors. This was attributed to an invalid present position used by MRU-A during alignment, possibly related to difficulties which arose during initial system turn-on. (The DEST DATA switch on the NCI was incorrectly set to M3 instead of B for initial MFCRS turn-on attempts.) Also, following the proficiency flight a large north/south (N/S) velocity was reported for MRU-A. This reported N/S velocity, without accompanying latitude, longitude, or East/West (E/W) velocity errors, was considered unlikely and this discrepancy was attributed to an incorrect octal number recorded from the NCI register after flight. (MFCRS Navigation data was displayed in octal on the NCI.)

Terminal position and velocity errors for MRU-A, and MRU-B were all better than basic F-15 INS specifications as shown by the position and velocity error plots in Figures 11 and 12 respectively.

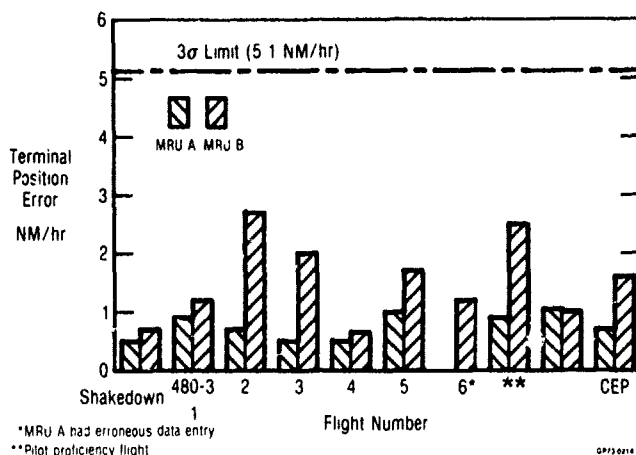


Figure 11. MFCRS Navigation Terminal Position Error

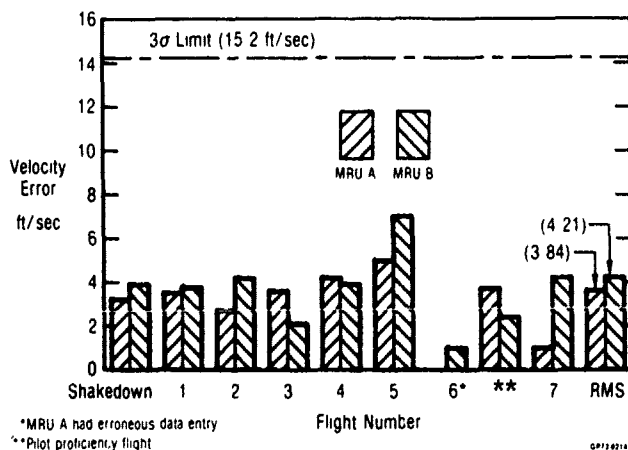


Figure 12. MFCRS Navigation Velocity Error

Conclusions From Phase 1 Flight Tests

Based on the results of the seven Phase 1 flights the following conclusions were obtained:

- Aircraft flying qualities with MFCRS sensors engaged were satisfactory at relatively low dynamic pressure (q) conditions (e.g., 0.6 Mach/30K Ft). However, aircraft damping with MFCRS sensors engaged at relatively high Q subsonic flight conditions (e.g., 0.9 Mach 15K Ft) was unacceptable following small aircraft perturbations.
- Acceptable aircraft behavior was noted following large amplitude maneuvers or rolling pull-ups with MFCRS sensors engaged, or following 1g to 3g MFCRS sensor disengagements at all conditions tested.
- MFCRS false alarm sensor faults observed during the first seven flights were eliminated by modification of the fault detection criteria.
- There were no MFCRS hardware failures during flight tests.
- MFCRS redundancy management logic properly detected simulated sensor faults.
- MFCRS generated valid navigation data that was better than the basic F-15 INS specifications.

Post Phase 1 Investigations

When the MFCRS low-damping problem appeared in the initial flight test results it became necessary to account for the differences between expected and actual system performance. A concerted effort was started to determine if any of the system components had been inaccurately modeled.

In order to isolate the problem area the components of the MFCRS and F-15 control system were divided into the categories shown in Figure 13. These categories were:

- Sensor components and switching unit - MFCRS and F-15
- F-15 CAS and Series Servo - From sensor output to series-servo output
- Actuator dynamics - rudder, stabilator
- Airframe Equations of Motion - rigid and elastic body.

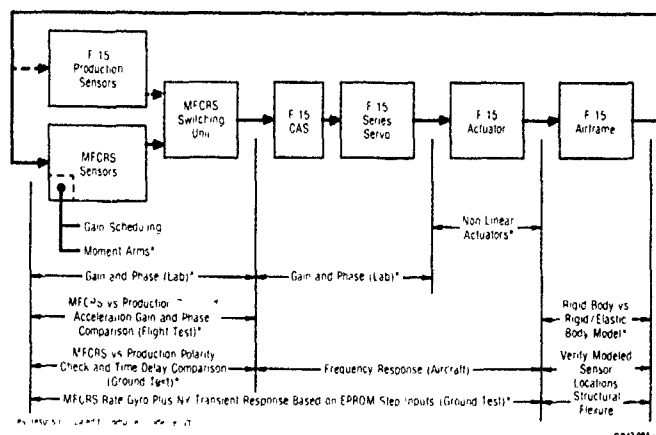


Figure 13. Low-Damping Problem Investigation

Sensor Analysis - Ground tests were conducted to assure that the expected magnitude, time delay, and polarity of the MFCRS signals out of the TMP were obtained. In these tests continuous recordings were made of each of the five flight control inputs to the CAS while the operator switched back-and-forth between MFCRS and STD. During this switching process the aircraft was manually moved in a fashion that provided sinusoidal outputs of pitch, yaw and roll. The results verified that the polarity of the MFCRS outputs were the same as production. Analysis of the strip charts also indicated that the gain and time delays obtained for the MFCRS and production sensors agreed closely with the analytic model.

Sensor gain and phase data for different input frequencies were obtained from 10° peak-to-peak rate table tests in the lab at St. Louis prior to the initial flight tests. The test results were in good agreement with the analytic model. After the initial flights (where the low damping in yaw was discovered), these rate table tests were repeated at EAFB using 0.25° to 1.0° peak-to-peak inputs. At this time the measurements indicated the yaw rate gain to be higher than expected at the 1 to 2 Hz input frequencies. This increase in the yaw rate gain was considered a contributor to the low-damping response experienced in the yaw axis.

CAS Plus Series Servo Analysis - Frequency response tests were performed in the MCAIR flight control laboratory on an F-15 Control Augmentation System including Series Servo (and excluding the actuator dynamics). The tests were performed to determine the accuracy of the "CAS plus Series Servo" model used in MFCRS analytical studies. Small amplitude outputs (gain and phase) were varied from 0.2 to 10 Hz. The results indicated excellent correlation between the lab data and the model except at the gain margin frequency of 2.1 Hz for the lateral axis. For this condition the rudder servo output in the lab was 1.3 dB higher than the model for a yaw rate input and 0.8 dB higher than the model for a lateral acceleration input. These results are such that they would contribute to a reduction in the damping for the yaw axis.

A test procedure defining tests to obtain frequency response data on the stabilator and rudder actuators was prepared and coordinated with AFFTC. These tests which were performed at Edwards Air Force Base (EAFB) applied sinusoidal commands at the pitch CAS input and measured the amplitude and phase characteristics of the unloaded actuator output motion at different frequencies. The tests were then repeated with the yaw CAS and rudder actuator. The results were analyzed and compared with similar data generated for the F-15 CAS, series servo, and actuator components of the analytical model, and with lab tests performed earlier at MCAIR.

The resultant data showed that the measured system gains at the critical gain margin frequency are generally higher by 1 to 2 db than the model gain. This higher gain is considered a potential contributor to the low damping observed in flight. The measured gains, however, closely match the gains of the model at the phase margin frequency thus the gain difference is due to more than just a scaling offset.

Rudder Actuator Analysis - A review of the rudder actuator receiving inspection frequency response data indicated the presence of nonlinear characteristics for small amplitude inputs. The model used in the MFCRS analysis was more characteristic of the rudder response at large amplitudes. The data showed an appreciable increase in phase lag is present at lower rudder-actuator input amplitudes. This increased lag was also considered a contributor to the low-damping problem experienced in yaw.

Stabilator Actuator Analysis - After the Phase I flights, ground test frequency response data was obtained to evaluate the stabilator actuator dynamics. A review of this data did not reveal a significant difference between it and the actuator model dynamics used in the MFCRS analysis. However, data obtained from an earlier F-15 Gust Alleviation Study (GAS) which performed an inflight frequency response on the stabilator actuators did show the presence of considerably more phase lag (at frequencies above 1 Hz) than obtained from the ground tests. These differences are attributed to the aerodynamic loading present on the actuator inflight. This was verified by a review of the actuator acceptance test results which showed increased lag when under load. A new stabilator actuator model was developed which more closely approximated the nonlinear actuator dynamics. When this model was incorporated into the analysis the pitch axis damping was more representative of that experienced in flight.

Airframe Dynamics - The airframe equations of motion used in the MFCRS analysis consist of three degree of freedom representations for the longitudinal and lateral-directional rigid body dynamics. Additional degrees of freedom were included for the following structural modes; first and second fuselage lateral bending, and first torsional bending. When the MFCRS low damping problem was uncovered in the initial flight tests, the approach and data used for the rigid/elastic body airframe representations were reviewed. It was determined that the aero data used in the MFCRS studies was consistent in detail, complexity and applicability for use in control system analysis of the MFCRS type.

Summarizing the above shows the following factors to have been the significant contributors to the low damping problem:

- Inflight stabilator actuator response had more gain and more lag than the model used in the MFCRS analytical model. This was verified by comparing the results from GAS flight tests (using modified actuators) and further tests on the F15 with standard actuators.
- Low amplitude rudder actuator response had more gain and more lag than the MFCRS model. This was verified by the results obtained during ground tests of the rudder actuator conducted on the aircraft.
- Yaw rate feed back output was higher than desired value (when rate was changing). This was verified by conducting laboratory calibration tests on the hardware.

- MFCRS pitch rate and yaw rate sensor outputs were significantly higher than production sensor outputs during pilot controlled frequency sweeps (especially in the 3-4 Hz range). This was verified by analysis of flight test results obtained during the MFCRS flights.

When the effects of these contributing factors were incorporated in the MFCRS analytical model the system compensation requirements were reevaluated. The analysis indicated that a TMP EPROM change could provide a significant improvement in the MFCRS yaw-axis damping and an improvement in pitch-axis damping. The following changes were implemented:

- o Yaw Axis
 - revise the Gain Scheduler Module
 - revise the scale factor of yaw-rate signal
 - add another 5dB lead-lag to yaw rate
- o Pitch Axis
 - revise the Gain Scheduler Module
 - replace twin-staggered notch with a dual-notch configuration
 - delete 15dB notch for normal acceleration

These performance improvements formed the basis for continuing the flight test evaluation of the MFCRS control system.

Phase II Flight Test

The specific objectives and success criteria for Phase II of the MFCRS Flight Test Program were:

Flight Test Objectives

- a) To verify that modified MFCRS is airworthy.
- b) Validate analytical design methods.
- c) To better define aircraft static bending characteristics. Validate bending model.
- d) To verify adequacy of moment-arm compensation.

Flight Test Success Criteria

- a) Positive damping in all three axes. No uncommanded motions occur.
- b) Flight test results approximate the nominal performance predicted for modified MFCRS.
- c) Flight test data shows expected bending amplitudes (inches /G) during LAM.
- d) MFCRS damping does not change when sensor trad $\Delta 123$ is used in lieu of $\Delta 456$.

Prior to the MFCRS Phase II Flight Test evaluation several tests were conducted in the MCAR lab facilities at EAFB to verify the following TMP EPROM changes:

- Addition of a second lead-lag filter and 15% gain reduction with respect to (WRT) nominal gain in the yaw-rate path.
- Modification of the 9 Hz and 14 Hz structural filters in the pitch-rate path.
- Expanding the gain scheduling function.

The tests included: 1) providing sinusoidal rate excitation to the system and monitoring corresponding rate outputs to verify proper operation and 2) spot checking the gain scheduler function at eight selected points (endpoints, center points and outside points) and comparing gain factor values to mathematically computed ones. The test results indicated the equipment was ready for installation in the test aircraft.

Functional ground tests were also performed on the test aircraft with MFCRS installed to verify proper system operation. The basic MFCRS operations, such as, align mode, fault mode, fault initiate, sensor selection and rapid disconnect were exercised. In addition, special TMP EPROMs were used to verify proper fault annunciations and corresponding surface movements upon fault insertions. A normal flight control check was made in both STD (Standard F-15 CAS sensors engaged) and MFC (MFCPC sensors engaged) modes to assure proper aircraft operation. Proper system operation was verified by these tests.

A second Structural Mode Integration (SMI) test was performed on the MFCRS equipment installed in the test aircraft at Edwards AFB in June 1984. This test was conducted to verify that structural motion pickup in the MFCRS sensor outputs would not affect the operation of the MFCRS/CAS control system. The test consisted of the excitation of the stabilator and rudder control surfaces with a slowly varying sinusoidal command over the frequency range of 0 to 25 Hz. The test evaluated the system characteristics using standard CAS sensors and MFCRS sensors from MRU-A (Triad 123), MRU-B (Triad 456) and for

a combination of MRU-A and MRU-B (Triad 246) outputs. The test results showed that the standard and MFCRS sensor configurations did not sustain or reinforce any structural mode oscillation. With these favorable SMi results, the go-ahead was given for the Phase II flight test evaluation.

The MFCRS test aircraft was flown twice in July 1984 to evaluate the Phase II system modifications. In these flights the pilot exercised both the MFCRS and STD F-15 control system with small amplitude maneuvers (loops, rolling pullups), and sinusoidal frequency sweeps. These inputs were applied systematically at altitudes of 7500, 15000, and 30000 feet and Mach No's ranging from 0.6 to 0.9.

The MFCRS flying qualities based on pilot comments showed that satisfactory aircraft damping was obtained with the MFCRS control system for small amplitude maneuvers at all flight conditions evaluated except at 0.9 Mach at 7500 feet and 15000 feet. At these conditions a noticeable degradation in aircraft damping was noted. The lower damping at these two high dynamic pressure flight conditions was predicted.

A commentary on the similarities and differences observed between the response of the MFCRS and STD F-15 control systems is presented in Figure 14. This commentary is based on the performance observed from the strip chart data obtained during the Phase II flights. A comparison of this data with the corresponding data obtained in the initial flight tests shows the considerable improvement in performance achieved with the MFCRS Phase II configuration.

Control Axis	Flight Condition	Commentary
Pitch Axis	Altitude 7,500 ft	
	Mach 0.6 (Figure 5.9a)	-- MFCRS Pitch Rate Response is Deadbeat With No Overshoot
	Mach 0.7 (Figure 5.9a)	-- MFCRS Pitch Rate Well Damped With Small Overshoot Present
	Mach 0.9 (Figure 5.10a)	-- Response of Standard F-15 Was Not Available at This Condition
		-- MFCRS Pitch Rate is Oscillatory ($\zeta=0.15$) Since Limitations on Lower Gain Scheduler Value Provides System With Only 2.5 dB Gain Margin. However This Damping is Better Than 0.1 Value Observed in February Flights
	Altitude 15,000 ft	
	Mach 0.6 (Figure 5.11a)	-- Correlation Between MFCRS and Standard F-15 Response is Good
	Mach 0.7 (Figure 5.12a)	-- The 4 Hz Oscillation on the Stabilizer Trace Relates to Longitudinal Stick Inputs. The CAS-On Response is Provided for Reference
	Mach 0.9 (Figures 5.13a, 5.14a, 5.15a)	-- The MFCRS Pitch Rate Response is Deadbeat With No Overshoot
		-- Standard F-15 Response Data Was Not Available at This Flight Condition
	Altitude 30,000 ft	
	Mach 0.6 (Figure 5.16a)	-- The Data Shows Excellent Correlation Between the MFCRS and Standard F-15 Responses
	Mach 0.7 (Figure 5.17a)	-- This Data Also Shows Excellent Correlation Between the MFCRS and Standard F-15 Responses
Roll Axis	Altitude 7,500 ft	
	Mach 0.6 (Figure 5.9a)	-- The Match Between the MFCRS and Standard F-15 Roll Response is Quite Good. The MFCRS Yaw Rate Output Contains a Component of the 8 Hz First Fuselage Lateral Bending Mode Which is Attenuated by Filters in the Standard CAS
	Mach 0.7 (Figure 5.9b)	
	Mach 0.9 (Figure 5.10b)	
	Altitude 15,000 ft	
	Mach 0.6 (Figure 5.11b)	-- The Roll Response of the MFCRS Compares Well With the Standard F-15 at These Conditions
	Mach 0.7 (Figure 5.12b)	
	Mach 0.8 (Figure 5.12b)	
	Mach 0.9 (Figures 5.13b, 14b, 15b)	
	Altitude 30,000 ft	
	Mach 0.6 (Figure 5.16b)	-- The Shape of the MFCRS Roll Response Compares Well With the Standard F-15. Although the MFCRS is Slightly Slower Than the Standard F-15 at This Altitude
	Mach 0.7 (Figure 5.17b)	
Yaw Axis	Altitude 7,500 ft	
	Mach 0.6 (Figure 5.9c)	-- There is Good Similarity Between the MFCRS and Standard F-15 Response to a Rudder Kick. The 8 Hz Fuselage First Lateral Bending Mode Response is Observable on the MFCRS Yaw Rate Gyro Output
	Mach 0.7 (Figure 5.9c)	
	Mach 0.9 (Figure 5.10c)	
	Altitude 15,000 ft	
	Mach 0.6 (Figure 5.11c)	-- The Frequency and Damping of the MFCRS Yaw Rate Responses Match That of the Standard F-15 Quite Well. The Low Damping Observed on the CAS-On Response is Representative of the Basic Airframe Response at This Altitude
	Mach 0.7 (Figure 5.12c)	
	Mach 0.9 (Figures 5.13c, 14c, 15c)	
	Altitude 30,000 ft	
	Mach 0.6 (Figure 5.16c)	-- The MFCRS and Standard F-15 Yaw Responses at These Flight Conditions are Also Quite Similar as Shown. Traces of the 8 Hz Fuselage First Lateral Bending Mode are Observable on the Yaw Rate Output
	Mach 0.7 (Figure 5.17c)	

Figure 14. Phase II Flight Test Response Commentary

The greatest improvement in system performance in Phase II was observed in the yaw-axis response to a rudder kick. The data shows the MFCRS and STD F-15 yaw response to be nearly identical at all of the test conditions flown. This is in contrast to the responses obtained in the February tests. The improvement in the MFCRS yaw-axis performance is attributed to the added yaw-rate lead-lag filter and the revised gain scheduler.

Considerable improvement was observed in the MFCRS pitch-axis response over that obtained in the initial tests. Even at the high Q flight condition (.9 Mach at 7500 ft), where the MFCRS gain margin is low due to restrictions placed on the gain scheduler, to minimize switching transients, the system exhibited better performance than was obtained in the initial tests. The MFCRS roll axis response for the small amplitude inputs is shown to be essentially the same as the STD F-15 roll response. This was also the case in the February flight tests. No changes were made to the MFCRS Phase II roll control system.

Conclusions from Phase II Flight Tests (July 1984) - The increase in the control system damping predicted for the Phase II flight tests was achieved. This was accomplished by revising the system filter compensation to offset the system lags (time delay) present in the rudder and stabilator actuators for low amplitude inputs and by revising the gain scheduler and the yaw output scale factor to reduce the system gain.

The pilot's evaluation of the MFCRS Phase II performance was positive as indicated by their following comments:

- The differences in response between the MFCRS and standard control system were minor and noticeable only because I was looking at them.
- The disengage transients between standard control system and MFCRS sensor output during a 3g pull-up were negligible.
- The predicted improvement in the modified MFCRS was there.

These pilot comments were substantiated by the recorded flight data in which favorable comparisons of the MFCRS and STD F-15 control system responses to stick raps and rudder kicks were obtained.

Enhanced MFCRS

Following the conclusion of the Phase II flight test program a definition study was conducted to determine what hardware and software changes could be made to the MFCRS to satisfy the original program goals. This study was performed using the improved aircraft model and new analysis techniques developed from data acquired during the flight test program. Improvements were defined and the subsequent analysis showed that with minor hardware changes and some software improvements the original goals of level 1 handling in subsonic and supersonic flight could be achieved. A summary of the major changes is:

- o Reduction of time delays:
 - Pitch Rate 54ms to 22ms
 - Normal Acceleration 146.3ms to 21.5ms
- o Scheduling of filter coefficients as a function of dynamic pressure (Q)
- o Reduction of acceleration channel noise by decreasing gyro quantization from 2 arc seconds to .5 arc seconds instead of the 100 ms filter.
- o Replacement of acceleration analog filters with digital filters in the 28000 preprocessor.
- o Design of filters using the in flight actuator model from the gust alleviation study.
- o Precalculation of filter coefficients resulting in a reduction of time delay in the filter outputs.
- o Reduction of the zero order hold delay in the sensor accumulator.
- o Removal of the 100ms filter in the acceleration channel due to reductions in acceleration channel noise by decreasing the gyro quantization from zero seconds to 0.5.

This new configuration was called the EMFCRS.

CONTROL LAW DEVELOPMENT - The EMFCRS flight control laws were developed using guidelines defined by the July 1984 ground and flight test results. An additional consideration was to reduce the system phase lags in an attempt to improve the aircraft handling qualities and stability characteristics. The hardware, software, and control law modifications significantly improved system stability and damping as shown in Figure 15.

The EMFCRS structural filter design was based on the results of the July 1984 structural mode interaction tests. Those results indicated that the primary problems were structural resonances at 9 and 14 Hz. These resonances were due to longitudinal modes. The aircraft lateral/directional modes were not a problem due to their lower amplitude and sufficient filtering in the yaw and roll axes. However, basing the new design on the July 1984 SMI results proved to be a mistake as the results were only for a fully fueled aircraft and we would encounter problems with a low fuel configuration.

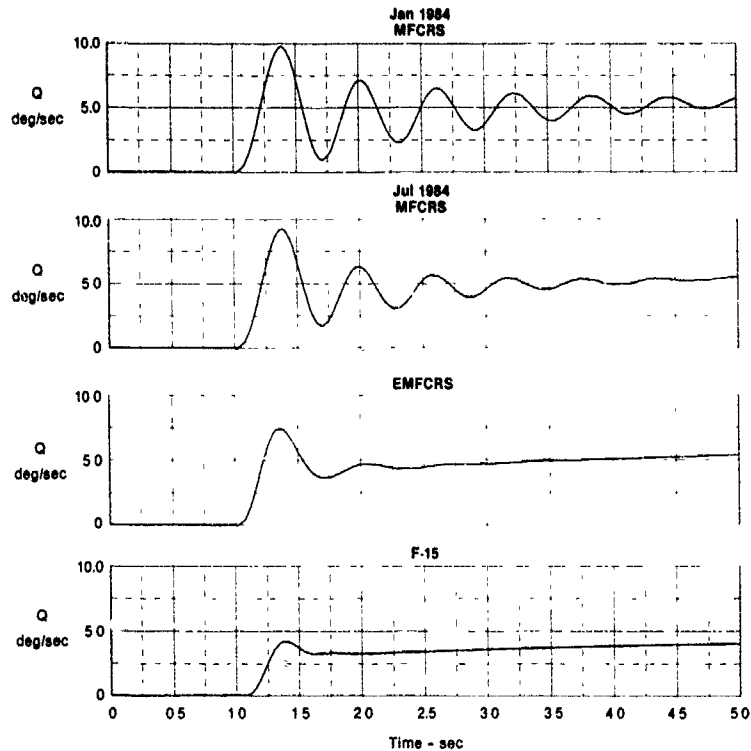


Figure 15. Evolution of Pitch Rate Response
Mach 0.9 at 7,500 Ft 10 Lb Long Stick

Two pitch axis configurations were defined: a primary system with reduced structural mode filtering for improved handling qualities, and a backup configuration with structural mode filtering similar to that flown in July of 1984. This precaution was taken because of uncertainties in the F-15 structural model; the backup configuration was selected based on the results of the SMI tests in July 1984 and gave us confidence since it had already been flown. Figure 16 shows the MFCRS stability envelope evolution.

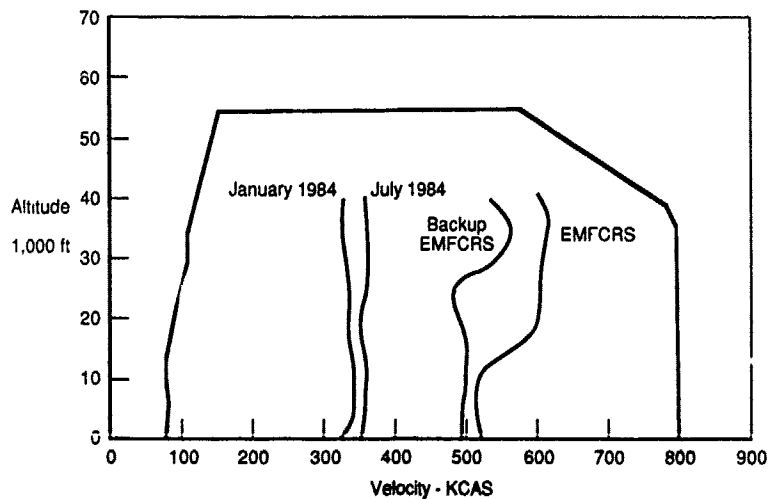


Figure 16. MFCRS Level I Stability Envelope Expansion

Figures 17 and 18 illustrate the changes to the MFCS rate and acceleration channels that resulted in the EMFCS configuration.

Figures 19 and 20 document the equivalent delay reductions achieved in the critical pitch channel for the EMFCS program.

REDUNDANCY MANAGEMENT (RM) - For EMFCS, the parity equation logic was updated to incorporate knowledge gained during follow on analyses and tests. The rates and accelerations used to calculate the trip level values were rescaled, to more closely duplicate the standard F-15 CAS disengage levels. In addition, a new structure was used for the accelerometer trip level calculations to account for coupling roll rate into the acceleration channel.

The updated EMFCS gyro and accelerometer trip level calculations are shown in TABLES 3 and 4 respectively.

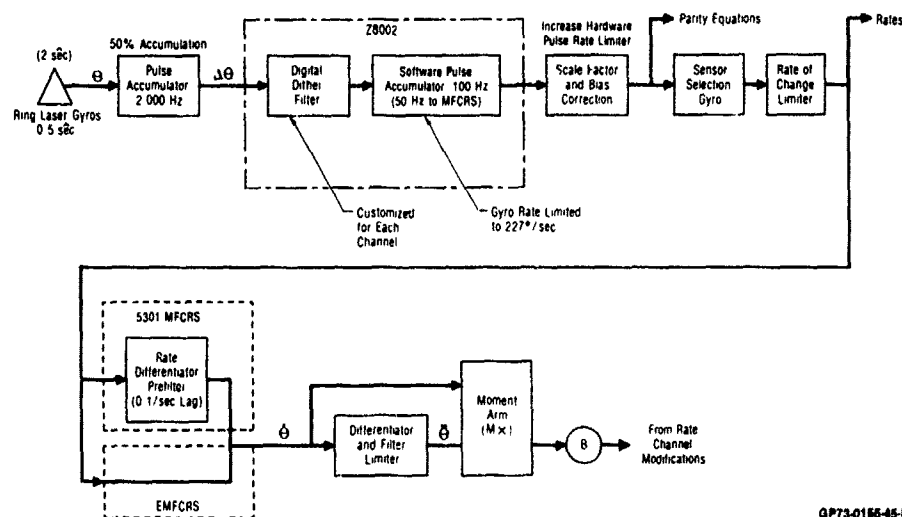


Figure 17. Rate Channel Modifications

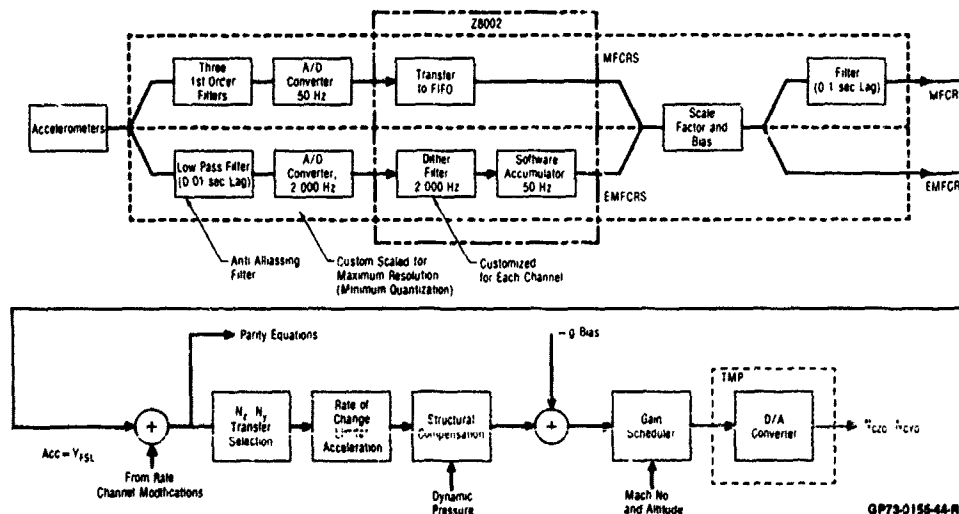


Figure 18. Acceleration Channel Modifications

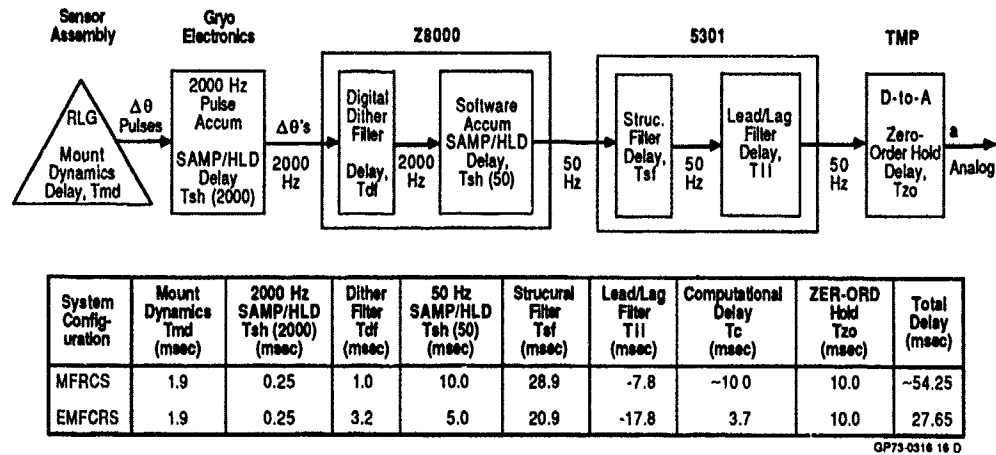


Figure 19. Pitch Rate Delay Components

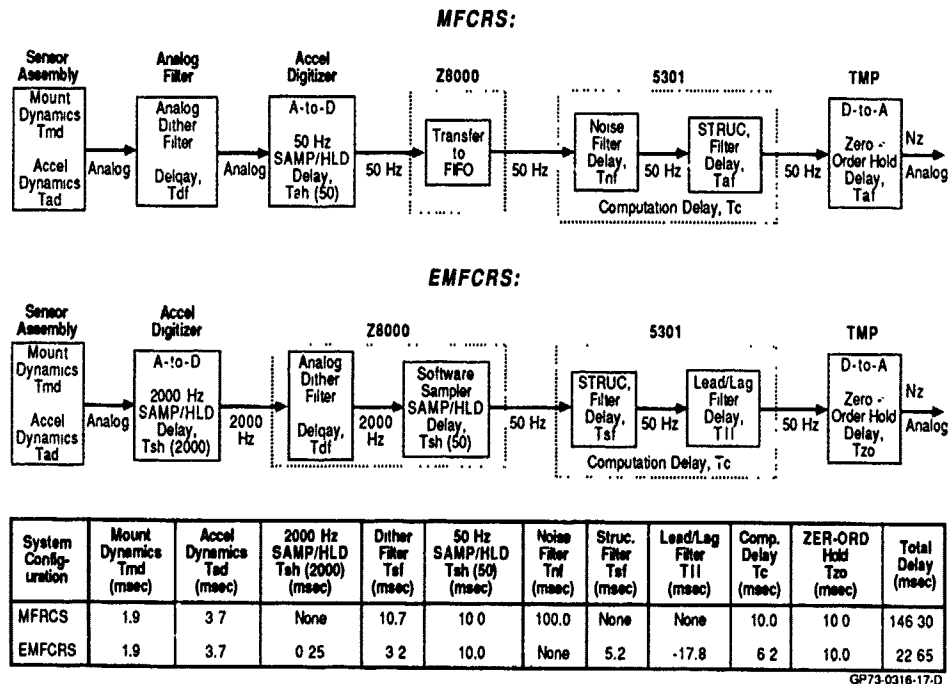


Figure 20. Normal Acceleration Delay Components

The peak-to-peak noise measurements on the actuators during integration testing showed a slight reduction in EMFCRS when compared to earlier testing of the original MFCRS as indicated in TABLE 5. These measurements indicate that significant noise components in the EMFCRS TMP outputs are in a frequency range outside the CAS/actuator response bandwidth. Therefore, the system noise reduction and system time lag improvements did not increase actuator noise and may have reduced it.

STRUCTURAL MODE INTERACTION TEST - The objective was to verify that EMFCRS compensation adequately attenuated structural mode interaction induced by vibration. To test the compensation, the stabilator and rudder control surfaces were driven with frequency sweeps from 0 to 25 Hz. In follow-on tests during November 1986, only the horizontal tail surfaces were excited. But the frequency sweeps were extended to 30 Hz.

TABLE 3. EMFCRS GYRO PARITY EQUATION TRIP LEVELS

$$\text{Fault Detection Trip Level} = (K14 \times \text{RATE1} + K23 \times \text{RATE2}) \times K24$$

or

$$= K13 \times K24 \text{ if } K13 > K14 \times \text{RATE1} + K23 \times \text{RATE2}$$

or

$$= K15 \times K24 \text{ if } K15 < K14 \times \text{RATE1} + K23 \times \text{RATE2}$$

$$\text{Sensor Selection Trip Level} = (\text{RATE2}) \times K24$$

or

$$= \text{Fault Detection Trip Level} \times K24 \text{ if } < \text{RATE2}$$

Where

RATE1 = P, K11 x Q, or K12 x R Whichever is Larger

$$\text{RATE2} = K17 \times P \times (P + K18 \times Q + K19 \times R + K20 \times \text{NZ}) + K21 \times \text{NZ} \times R + K22$$

P = Roll Rate

Q = Pitch Rate

R = Yaw Rate

NZ = Normal Acceleration

Constants	EMFCRS Value	July 1984	Derivation for EMFCRS Value
K11 =	10	43	P Maximum, Q Maximum or R Maximum
K12 =	10	1.5	
K13 =	0 078 rad/sec	0 078	112% of Minimum CAS Trip Level for Q&R
K14 =	0 07	0 07	7%
K15 =	0 366 rad/sec	0 366	Maximum CAS Trip Level for P
K17 =	2.5×10^{-3}	2.5×10^{-3}	
K18 =	89	89	
K19 =	89	89	
K20 =	2.62×10^{-2}	2.62×10^{-2}	
K21 =	6.52×10^{-5}	6.52×10^{-5}	
K22 =	0 044 rad/sec	0 052	85% of Minimum CAS Trip Level for Q&R
K23 =	0 07	0 01	7%
K24 =	0 2357		PE Normalization Factor = $1/\sqrt{18}$

Note: P, Q, R, fault detection trip level and sensor selection trip level are in rad/sec. NZ is in ft/sec².

GP73-0156-28-R

TABLE 4. EMFCRS ACCELEROMETER PARITY EQUATION TRIP LEVELS

$$\text{Fault Detection Trip Level} = (K3 \times \text{ACCEL1} + K25 \times \text{RATE1}) \times K24$$

or

$$= K2 \times K24 \text{ if } K2 > K3 \times \text{ACCEL1} + K25 \times \text{RATE1}$$

or

$$= K4 \times K24 \text{ if } K4 < K3 \times \text{ACCEL1} + K25 \times \text{RATE1}$$

and

$$\text{Sensor Selection Trip Level} = \text{ACCEL2}$$

or

$$= \text{Fault Detection Trip Level} / K24 \text{ if } < \text{ACCEL2}$$

Where

$$\text{ACCEL1} = \text{NZ or } K1 \times \text{NY if } K1 \times \text{NY} > \text{NZ}$$

$$\text{ACCEL2} = K26 \times \text{RATE1} + K9$$

P = Roll Rate (in rad/sec)

NY = Lateral Acceleration

NZ = Normal Acceleration

RATE1 = Primary Rate (in rad/sec)

Constants	EMFCRS Values	July 1984	Derivation for EMFCRS Value
K1 =	10	45	Maximum NZ or Maximum NY
K2 =	5 25 ft/sec ²	5 25	150% of Minimum CAS Trip Level for NY
K3 =	0 10	0 05	10%
K4 =	30 0 ft/sec ²	30 0	Maximum CAS Trip Level for NZ
K6 =	N/A	0 022	
K7 =	N/A	0 00296	
K8 =	N/A	0 0245	
K9 =	3 5 ft/sec ²	3 5	Minimum CAS Trip Level for NY
K10 =	N/A	0 01	
K24 =	0 2357	0 2357	PE Normalization Factor = $1/\sqrt{18}$
K25* =	1 8		Roll Rate Coupling into Acceleration Channel
K26** =	0 8		Roll Rate Coupling into Acceleration Channel

Notes

1 NY, NZ, fault detection trip level and sensor selection trip level are in ft/sec².

P is in rad/sec

2 For MFCRS

Fault detection trip level = $(K3 \times \text{ACCEL1} + K10 \times \text{ACCEL2}) \times K24$ ACCEL2 = $K6 \times \text{NZ} \times (P + K7 \times \text{NZ}) + K8 \times \text{NY} \times P + K9$

New constant, K8 used in actual software implementation

**New constant, K8 used in actual software implementation

GP73-0156-27-R

**TABLE 5. STATIC NOISE ON ACTUATORS -
MFCRS vs EMFCRS**

Triad	Peak-to-Peak Motion (deg)			
	Rudder		Stabilator	
	MFCRS	EMFCRS	MFCRS	EMFCRS
456	0.03	0.03	0.05	0.03
246	0.13	0.12	0.18	0.06

GP73-0316-12 R

The aircraft was checked for resonant frequencies with standard F-15 sensors and again with EMFCRS sensors. The tests were initially run with partially fueled and later with fully fueled airframe configurations.

Unacceptable aircraft oscillations occurred at 22 Hz during horizontal tail excitation (September 1986) for 0 to 25 Hz sweeps for a partially fueled aircraft when certain MRU sensor triads were used. The oscillations during SMI could not be explained, so the EMFCRS flight tests were cancelled and an investigation was begun. Post-test analyses, which related configuration, procedures, and data of the previous MFCRS SMI in July 1984, indicated the 22 Hz airframe oscillation could have been present, but not detected in July 1984 because the tests in July were with a fully fueled aircraft. Follow-on tests were successfully concluded with no 22 Hz oscillations during November 1986 with the aircraft fully fueled for a horizontal tail excitation from 0 to 30 Hz, confirming the July 1984 testing hypothesis. No problems were encountered during rudder excitation sweeps from 0 to 25 Hz.

STRUCTURAL SURVEY - The concept of the structural survey was conceived during the EMFCRS testing in August. The survey on F-15 S/N-77-139 was performed in November of 1986. The intent of the survey was to collect structural data in order to correctly model the F-15 airframe for use in the design of structural filters. The data was obtained by sweeping the various control surfaces while monitoring the sensor outputs. All data was gathered open loop, which means that the sensor outputs were decoupled from the CAS.

Data were recorded for the standard F-15 sensors, MFCRS MRU-A and MRU-B. Both partially fueled and fully fueled aircraft configurations were tested. Typical survey data and its corresponding math model are compared in Figure 21. The model was obtained by performing least squares curve fit on the frequency response data. The accuracy of the model was verified by analytically coupling it to the EMFCRS flight control system model. The MRU-A, low weight model was used since conditions of stability and instability were observed as the N_z signal was coupled and decoupled when locked on to MRU-A.

As Figures 22 and 23 illustrate, the model predicted stability with N_z coupled and predicted a 22 Hz instability with N_z decoupled respectively. Thus the model accurately predicted the EMFCRS behavior that was observed at EAFB.

Due to the fact that the ABICS III program was already in place and resolved many of the EMFCRS program hardware limitations, no attempt to correct the structural filtering in the EMFCRS system was made.

MFCRS Program Conclusions

The MFCRS Program has shown that it is feasible to use dispersed skewed inertial navigation quality sensors for redundant flight control sensors, navigation, weapon delivery, cockpit displays and sensor stabilization. The program has also shown that there are some additional requirements when designing a control system using dispersed sensors installed at nonoptimal locations in the aircraft. These requirements are:

1. Time delays are critical and should be kept to about 20ms in both the rate and acceleration channels.
2. Quantization levels of the rate sensor outputs should be less than 0.5 arc seconds.
3. Aircraft dynamics models must be well defined.
4. Actuator dynamics must be well defined for all flight conditions.
5. The design must be able to include the forward and feedback loop of the control system.

6. Dither noise on the sensor outputs must be minimized in the initial design of the sensor package to avoid large amounts of filtering of these outputs.
7. A total digital design is preferable to avoid A/D and D/A conversions.
8. A sample rate of at least 80Hz is necessary to achieve the necessary bandwidth for the flight control system.
9. Full system integration tests in the laboratory are essential to the development and testing of the system.

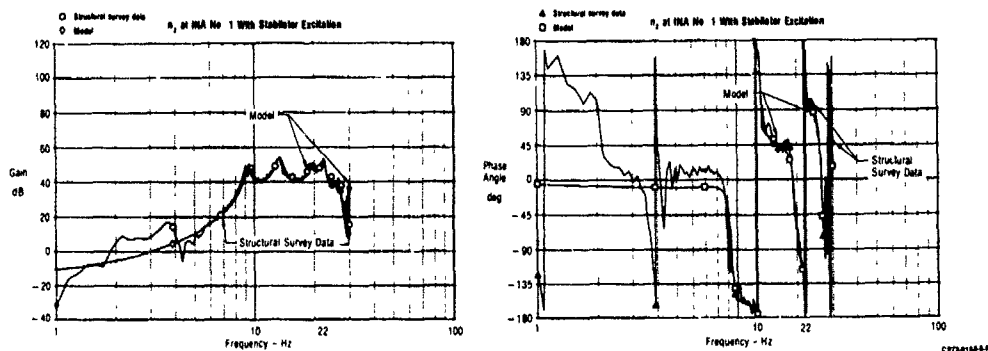


Figure 21. Gain Phase Comparison: Model vs Structural Survey Data

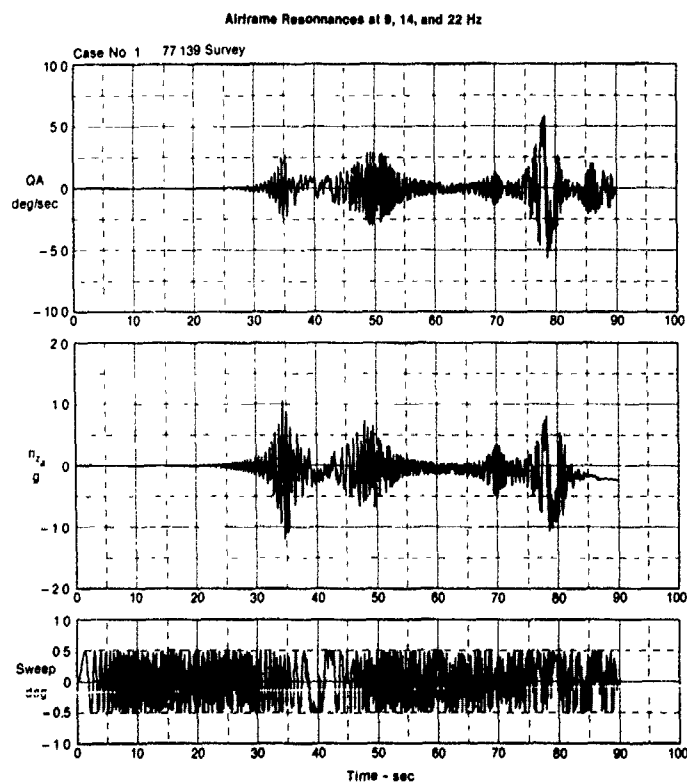


Figure 22. Simulated SMI MRU-A n_z Coupled

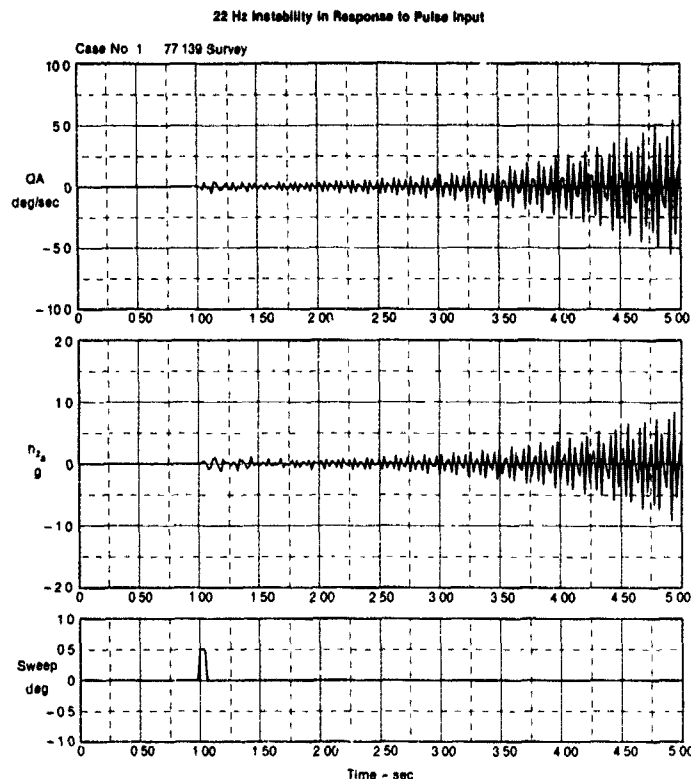


Figure 23. Simulated SMI MRU-A n_2 Decoupled

THE ABICS III PROGRAM

MCAIR has been involved in evaluating the Ada HOL in real time embedded systems applications programs since 1983/84 (Figure 24). At the same time it became evident to the Air Force at the conclusion of the MFCRS Program (1985/86), that to fully evaluate the multifunction concept, specially designed hardware and a digital flight control computer were required. At the same time, the NAVY was seeking a cost effective approach to test its IISA hardware. As a result, the joint ABICS III program emerged. The IISA hardware had digital outputs and the sensor block had been designed for navigation and flight control, it had reduced sensor quantization and fast (8 MHz) Z8002 micro-processors. The ABICS test aircraft had a digital flight control system with mechanical backup and an Integrated Flight and Fire Control (IFFC) system already on board. The modification of the digital flight control I/O to add the required interface to the IISA sensors was the only major hardware change required.

To take the next step in the evaluation of embedded systems using Ada, the navigation algorithms and the redundancy management would be programmed and flight tested. The test aircraft would now have Ada software to implement the flight control laws, IFFC, redundancy management of the flight control sensors and the dual inertial navigation system. A 5 channel Global Positioning System would also be on board to help score the navigation performance. Figure 25 shows the aircraft configuration. In September 1986 MCAIR received a contract from the Air Force and Navy to design and flight test the flight control system using the IISA sensors.

Figure 26 shows the functional block diagram of the ABICS III configuration. The goals of the ABICS III program are:

- o To demonstrate through flight test that navigation quality sensors in a strapdown configuration can be used as fault tolerant flight control references
- o To obtain real world experience using Ada in embedded integrated control systems
- o To establish confidence in the viability of Ada
- o To determine requirements for software engineering environments
- o To obtain metrics of Ada usage

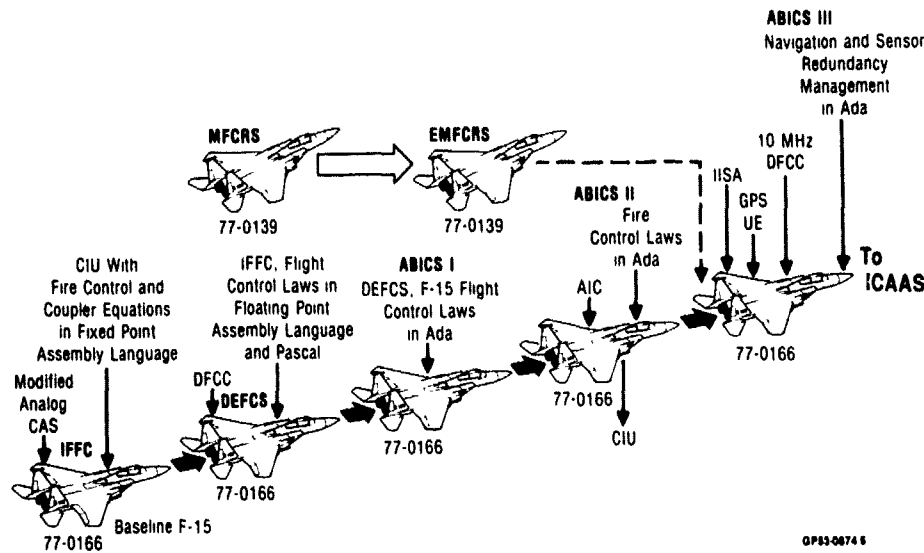
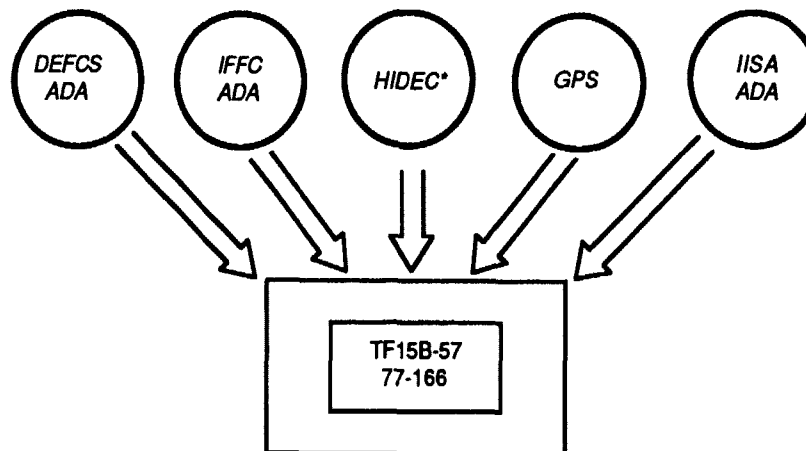


Figure 24. Ada Based Integrated Control System (ABICS)



*CONFIGURATION STUDIES ONLY

GP73-0214-34-D

Figure 25. ABICS III Aircraft Configuration

Figure 27 shows the aircraft modifications performed to install the ABICS III equipment and Figure 28 shows the hardware modifications for the program. In order to accommodate the added computational burden in the digital flight control computer, the Z8002 processor was upgraded to a 10 MHz version and a 10 MHz clock was implemented (from a 6 MHz version) thus increasing the throughput by 1.7 times. The ABICS II program had shown the need of a 32 bit architecture for the Ada version of IFFC so arrangements were made to host IFFC in a 32 bit processor for ABICS III by installing a Rolm Hawk/32 and upgrading the Avionics Integration Computer (AIC) to a 32 bit Zilog Z80K microprocessor.

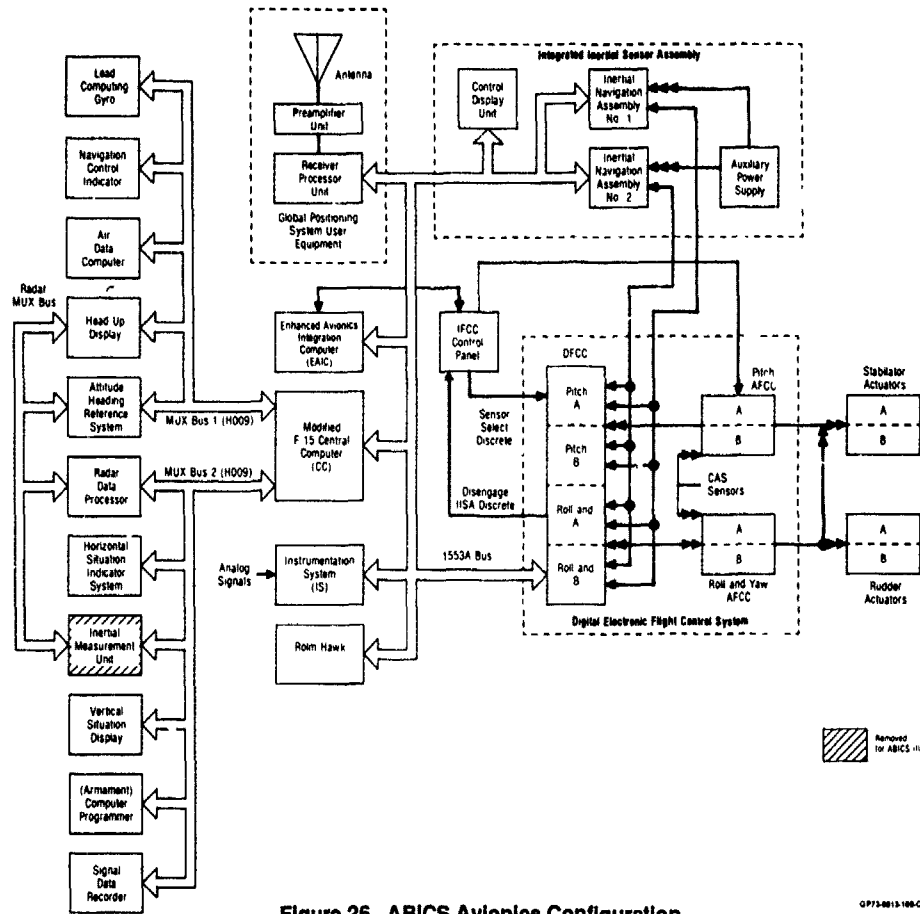


Figure 26. ABICS Avionics Configuration

QP73-0013-1000-0

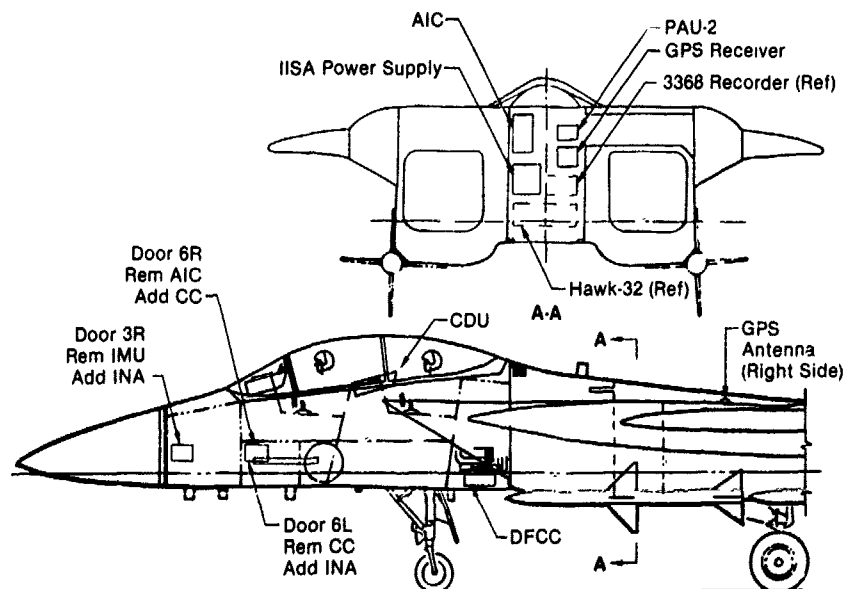


Figure 27. ABICS III Equipment Location

QP73-0013-1000-0

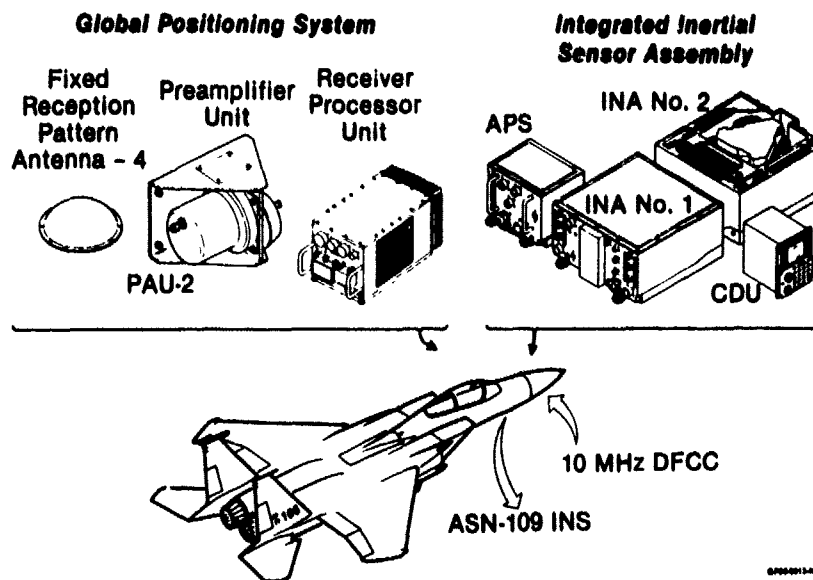


Figure 28. ABICS III Hardware Modifications

ABICS III Control System Design - In contrast to the MFCS program, the control system for ABICS III uses a digital flight control computer and allows total system design. Compensation can now be added in the forward or feed back loops. In order to ensure an aircraft structural model of high fidelity, a structural survey was conducted as in the EMFCS program on the ABICS III aircraft (a TF-15B, S/N 77-166). An interesting result of this work was that this aircraft did not show the 22 hz resonance seen on F15A 77-139. No explanation has been found. Surveys of other aircraft F15A's, B's, C's and E's would be of interest to characterize model dependent modes. Present analysis shows that the ABICS III control system will have the same stability envelope as the standard F15 and achieve the Multifunction Inertial Reference goals of providing equivalent flight control system response as the currently practice of using dedicated, co-located sensors mounted on nodes and anti-nodes. Figure 29 shows how closely the F15/IISA response will be to F15/standard sensors.

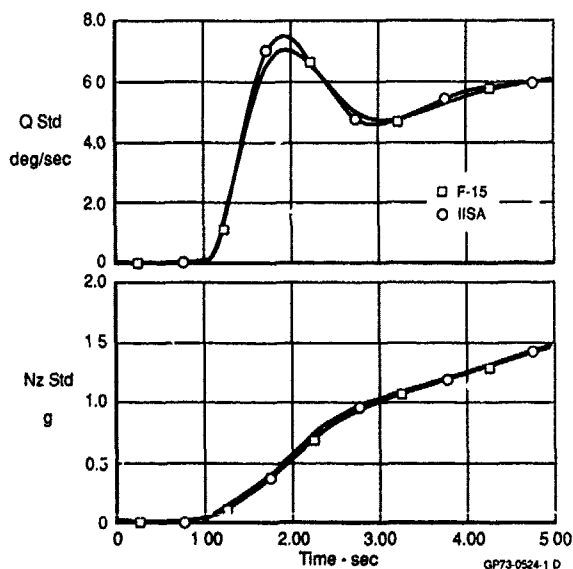


Figure 29. Response to 10 lb Long Stick Mech at 30k'

ABICS III Redundancy Management - The IISA redundancy management approach is somewhat different than the MFCRS. Although parity equations are used, the IISA RM uses a real time algorithm and not a table look-up.

The sensor geometry and methods of redundancy management are similar to those described in the literature. The sequence of operations performed in IISA is illustrated in Figure 30. Sensor data is first reviewed for hard failures, detectable by normal self-test methods. The sensors themselves give an indication of failures through loop closure tests, loss-of-signal indications, etc. I/O tests assure that data has been correctly transmitted, and dynamic reasonableness tests detect spurious outputs inconsistent with the vehicle capability.

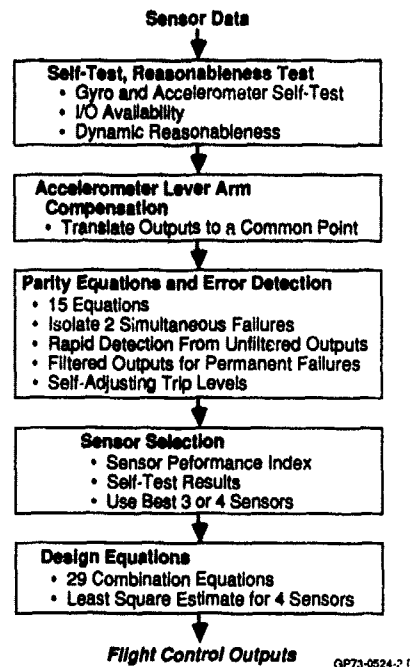


Figure 30. Redundancy Management Operations

Due to the physical separation of the two sets of accelerometers, angular rotations and angular accelerations of the vehicle cause different accelerations to be sensed by each set. To allow direct comparison between acceleration measurements under dynamic conditions, each sensor output is related to a common point on the aircraft using the current best estimate of vehicle angular rate and angular acceleration along with known lever arm displacements from that point.

The six skewed gyro axes and six skewed accelerometers are spaced evenly on a 109.5° cone whose axis is vertical. Since no two axes are coincident and no three are in the same plane, full three-axis outputs can be provided with three failures of a sensor type. Reasonable geometry is available for any combination of failures, i.e., geometrical amplification of errors is less than a factor of 2.3.

Detection of up to three failures is assured by comparison of redundant sensor data in what are termed parity equations. These equations cancel vehicle angular rate, or acceleration in the case of accelerometers, and expose sensor errors. Because of information limitations, a third sensor failure of the same type can only be detected. Isolation of which of the four sensors active at that point has failed and cannot be achieved except for hard failures which are detected by conventional self-test methods. For this reason IISA is termed fail-operational/fail-operational/fail-safe.

Six gyro (similarly for accelerometers) parity equations can be formed by comparing each gyro output to a least-squares estimate of its output derived from the remaining sensors. Since there are always two sensors orthogonal to each axis, this results in six equations which are linear combinations of four sensor outputs. The orthogonal sensors cannot contribute to error detection. After sensor failures, a different set of parity equations is required. Again, linear equations involving four sensors can be formed, five equations after the first failure and only one after the second failure.

GP73-0214 36-D

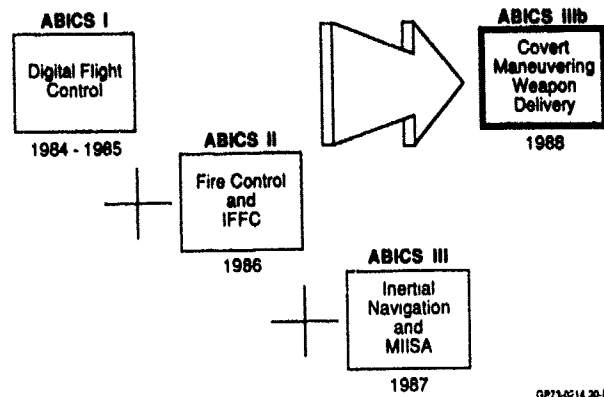


Figure 32. ABICS Programs
Applications of Ada

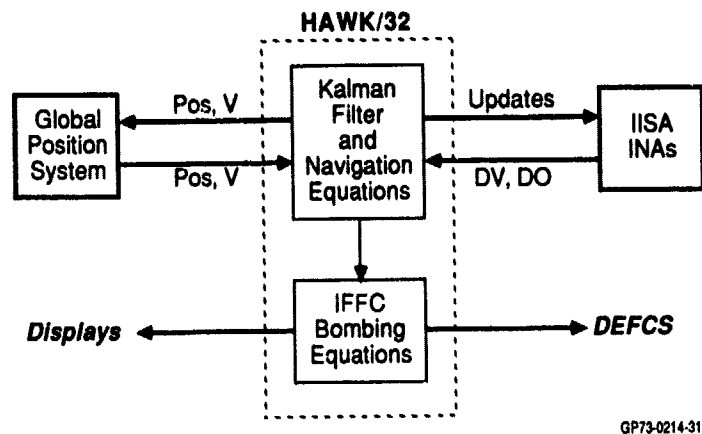


Figure 33. ABICS/32

Conclusions

The experience gained in multifunction sensor technology in the past 13 years has made it obvious that new techniques are required to design these systems, and that the structural dynamic models and the aeroelastic coupling to the sensors must be well defined for the target aircraft. Present multifunction sensors must also be reduced in size and weight while maintaining the same performance in order to stay within physical constraints of modern aircraft and allow dual installations. A thirty pound per sensor box weight goal would satisfy most applications.

References

1. Perdsock, J. Air Force Flight Dynamics Laboratory; Burns, R. C., McDonnell Douglas Corporation - "Preliminary Feasibility Assessment of Multifunction Inertial Reference Assembly (MIRA)", presented to American Defense Preparedness Association, Avionics Sections, Air Armaments Division Technical Symposium 4-5 October 1978 at Naval Surface Weapons Center, White Oak Laboratory.
2. Sebring, D. L., McDonnell Douglas Corporation; Young, Captain, J. T., Air Force Wright Aeronautical Laboratories, Flight Dynamics Laboratory - "Redundancy Management of Skewed and Dispersed Inertial Sensors", AIAA Paper Nr 81-2296, presented at the Fourth Digital Avionics Systems Conference, 17 - 19 November 1981, St. Louis, MO.

3. Luedde, W. J., McDonnell Douglas Corporation - "The Use of Separate Multifunction Inertial Sensors for Flight Control", AIAA Paper Nr 81-2295 presented at the Fourth Digital Avionic System Conference 17 - 19 November 1981, St. Louis, MO.
4. Young, Captain, J. T., Perdsock, J., Air Force Flight Dynamics Laboratory, Edinger, D. L., McDonnell Douglas Corporation; Edinger, L. D., Honeywell Inc., - "Design and Development of the Multifunction Flight Control Reference System", presented to advisory group for Aerospace Research and Development (AGARD), Guidance and Control Panel Technical Symposium, Enase, Toulouse France 17 - 20 May 1983.
5. Sebring, D. L.; Young, J. T.; and Perdsock, J.; "Application of Multifunction Inertial Reference System", AGARD Lecture Series 133, May 1984.
6. Palmer, R. H., Schwent, D. R., and Segnfredo, G. A., McDonnell Douglas Corporation - "Laboratory Evaluation of Strapdown Multifunction Flight Control Reference System", presented at Joint Services Data Exchange for Inertial Systems, October 1984, Salt Lake City, Utah.
7. Bedoya, C. A., et al., McDonnell Douglas Corporation - "Overview of Multifunction Flight Control Reference System Development," presented at the 12th Biennial Guidance Test Symposium, October 1985, Holloman AFB, N.M.
8. Bedoya, C. A., McDonnell Douglas Corp; Perdsock, J. M. - "Overview of the IISA/ABICS Flight Test Program, ION Paper presented at the 43rd Annual Meeting of the Institute of Navigation, June 1987, Dayton, Ohio.
9. Ebner, R. E., Litton Systems Inc. - "Integrated Inertial Sensor Assembly Advanced Development Model," Presented at the Seventeenth Joint Services Data Exchange for Inertial Systems, 15-18 October 1984.

KALMAN FILTER FORMULATIONS FOR TRANSFER ALIGNMENT OF STRAPDOWN INERTIAL UNITS

by

Alan M. Schneider
University of California, San Diego
Department of Applied Mechanics & Engineering Sciences
La Jolla, California 92093
United States

ABSTRACT

Formulations of Kalman filters are presented which are capable of aligning one strapdown inertial sensor assembly with another by estimating the misalignment angle between them. One formulation treats the case of a fixed misalignment. Another treats the case of a dynamic misalignment, caused, say, by bending of the common supporting body. Measurements can be made by gyros only, or by gyros plus accelerometers. Filters which estimate inertial sensor error parameters are also discussed.

THE PROBLEM

Consider the problem of estimating the small mechanical misalignment between the case axes of two strapdown inertial sensor assemblies (ISA's) mounted on a common body and separated by a sizable distance. For example, one ISA could be in the cockpit of a fighter aircraft, and the other could be in a missile on the wing. It is presumed that the separation distance is such that, with intervening structure, it is not possible to use optical alignment methods. It is also presumed that mechanical misalignments between the two locations may be large compared to the alignment accuracy required, so that simply referencing the cases to the frame of the structure is not adequate. This problem is sometimes called the "transfer alignment" problem.

The first situation considered is that in which the body is rigid, the ISA's have gyros only (no accelerometers), and no attempt is made to estimate the error parameters of the inertial sensors. Later, we consider: a) the effects of a nonrigid body, b) adding accelerometers to the ISA's, and c) estimating error parameters of the inertial sensors, such as gyro drifts and accelerometer biases.

The angular motion of the body excites the angular velocity sensors (i.e. the gyros on each ISA). Observation of the response of each ISA to the common angular velocity provides information for estimating the misalignment, assumed to be small, between the coordinate frames defined by the case axes of each unit. If the misalignment can be estimated, then a mathematical correction can be made to the output of ISA2, so that it is effectively "aligned."

PRIOR WORK

Reference 1 presents a least-squares solution to the problem of estimating the (3×3) transformation matrix between the two coordinate systems. This approach does not lend itself well to real-time, on-line implementation. Reference 2 presents a 36-state Kalman filter for this application. While a Kalman filter in principle is well adapted to real-time implementation, a 36-element state vector is much larger than necessary and could be difficult to implement, especially in a typical airborne computer. Also, it is so large as to obscure to the non-specialist what is taking place in the alignment process. Reference 3 discusses a Kalman filter used to align an inertial platform in a SRAM missile carried on the wing of an airplane, relative to another inertial platform used to navigate the aircraft. However, the technique is based on position-matching (the position being that derived from each inertial navigator), and is not applicable in the problem posed here, since without accelerometers, there is no capability on the part of the ISA's to compute position. In addition, the difference between the stabilized platform and strapdown is significant.

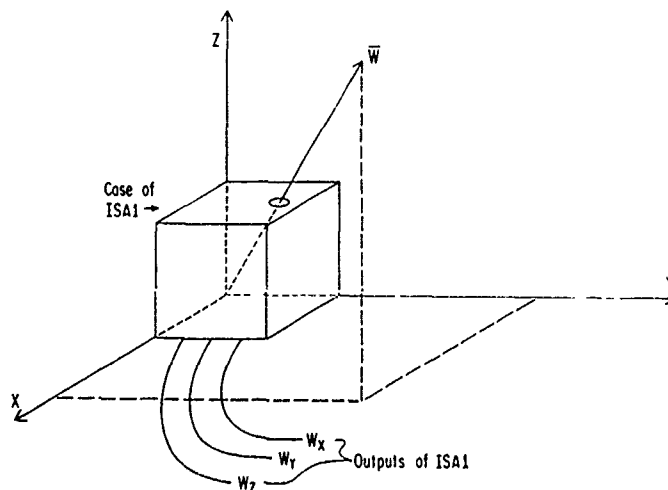


Fig. 1. -The ISA as an Angular-velocity sensor.

APPROACH

For present purposes, it is convenient to think of a strapdown ISA not as an inertial navigator or attitude reference system, but simply as a convenient sensor for measuring the vector angular velocity of its case with respect to inertial space. The output of an ISA is therefore taken as the three components of the angular velocity vector resolved onto the case axes. See Fig. 1. When the supporting body is rigid, rotational motions of the body, such as roll, pitch, yaw, or purposeful turning maneuvers, are sensed and measured by both ISA's. The difference in the output of the two ISA's provides the basic information which can be used by a filter to estimate the misalignment. If, for example, the two ISA's were perfectly aligned, and had perfect gyros, then the set of outputs of ISA1 would equal the set of outputs of ISA2. On the other hand, small differences in the outputs, axis by axis, are indicative of misalignment.

It turns out, that when the estimation problem is properly formulated, the misalignment angle is linearly related to the difference in the angular velocity output of the two ISA's. Thus, the Kalman filter, a means of linear estimation theory, is ideally suited to the task. Therefore a Kalman filter will be used as the primary tool for carrying out the alignment, that is, it will be used to estimate the misalignment angle of ISA2 relative to ISA1, as shown in Fig. 2.

The assumption that the misalignment is constant corresponds physically to the ISA cases being rigidly attached to the body, the body itself being rigid, the inertial instruments inside being rigidly attached to the case, and the entire collection being free of dimensional and performance changes with temperature or load or other disturbing influences.

Let the first ISA be represented by its case axes x, y, z , referred to as the "fixed" frame F , with unit vectors $\hat{i}, \hat{j}, \hat{k}$, and the second be represented by its case axes x, y, z , called the "moving" frame M . Both F and M are assumed to be right-handed Cartesian frames. Let $\vec{\phi} = \hat{i}\phi_x + \hat{j}\phi_y + \hat{k}\phi_z$ be the small, unknown, constant misalignment angle of M relative to F . It is this vector which is to be determined. The situation is pictured in Figure 3. Here the F and M frames are shown, misaligned by a small angle, and both subjected to the angular velocity \vec{W} . (For simplicity, only two dimensions are illustrated.) One measurement will consist of the output of the three gyros on each ISA. On the first ISA, the output is $[W_x, W_y, W_z]^T$, the three F -frame components of \vec{W} . On the second ISA, the output is $[W_x, W_y, W_z]^T$, the three M -frame components of \vec{W} .

In practice, when integrating gyros are used, the "measurement" consists of integrating \vec{W} over a short interval Δt , with the three outputs of the first ISA, for example, being increments of angle $[W_x\Delta t, W_y\Delta t, W_z\Delta t]^T$, available at the end of the measurement interval. As long as the start and end of the measurement interval is the same for all three gyros on both ISA's and as long as Δt is short enough so that the angular velocity vector does not change direction appreciably, the analytical results to be derived below are essentially the same as if \vec{W} is considered to be the physical quantity that is measured. One would make slight adjustments in the present formulation to handle the case of measuring increments of angle.

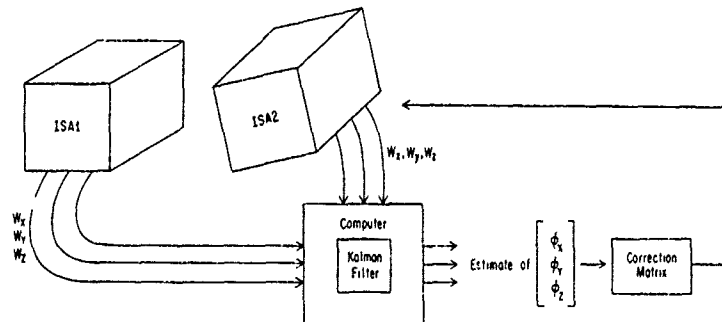


Fig. 2. -Correcting alignment with a Kalman filter.

Because of the misalignment, the components of \vec{W} in the F frame, one by one, will not equal the components of \vec{W} in the M frame. Thus, even though both ISA's measure the same vector \vec{W} , they come to different results. Considered as a 3×1 column matrix, the output matrix of ISA1 will not equal the output matrix of ISA2,

$$[W_x, W_y, W_z]^T \neq [W_x, W_y, W_z]^T \quad (1)$$

It is this inequality which provides the hope for deciphering the magnitude and direction of $\vec{\phi}$.

Another way of interpreting the difference between the matrix of outputs of the two ISA's is to think of the ISA's as being aligned, but measuring two different angular velocities. The first of these two angular velocities is the original \vec{W} , but the second, \vec{W}' , is the original \vec{W} rotated by $(-\vec{\phi})$. \vec{W} is measured by ISA1; \vec{W}' is measured by ISA2. The vectors measured in this hypothetical case are illustrated in Figure 4. The outputs $[W_x, W_y, W_z]^T$ and $[W'_x, W'_y, W'_z]^T$ will be exactly the same as in the real situation in Figure 3. From Figure 4, it is evident that the following vector cross-product relationship holds (provided that the magnitude ϕ of the vector $\vec{\phi}$ is small enough so that its sine is essentially equal to the angle)

$$d\vec{W} = \vec{\phi} \times \vec{W} \quad (2)$$

where

$$d\bar{W} = \bar{W} - \bar{W}' \quad (3)$$

The three F-frame components of $d\bar{W}$, denoted by dW_x, dW_y, dW_z , arranged as a 3×1 matrix, equals the difference between the output matrices of the two ISA's.

$$\begin{bmatrix} dW_x \\ dW_y \\ dW_z \end{bmatrix} = \begin{bmatrix} W_x \\ W_y \\ W_z \end{bmatrix} - \begin{bmatrix} W'_x \\ W'_y \\ W'_z \end{bmatrix} = \begin{bmatrix} W_x - W'_x \\ W_y - W'_y \\ W_z - W'_z \end{bmatrix} \quad (4)$$

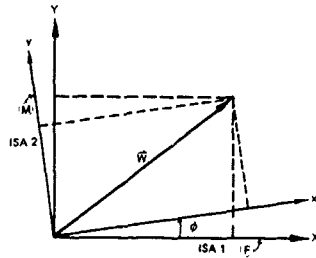


Fig. 3. -ISA measurements of angular velocity: one vector in space measured from two coordinate frames.

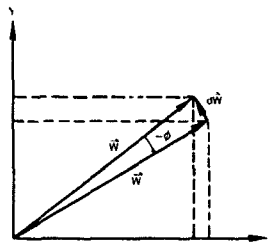


Fig. 4. -Alternative interpretation: two vectors in space measured from one coordinate frame.

KALMAN FILTER FORMULATION

Because we will get measurements from time to time (say, every Δt seconds), we use the discrete form of the Kalman filter. The filter and notation are that of Ref. 4, p 110. Table I, taken from the reference, displays the standard form of the discrete Kalman filter equations. The vector x_k is the $(n \times 1)$ state vector of the filter at time t_k , and is the quantity to be estimated. The state vector propagates from time t_k to time t_{k+1} through the dynamics of the process, as embodied in the state-transition matrix Φ_k . In the general case, white noise w_k , a random sequence referred to as "process" noise, also drives the state vector in an unpredictable way. The white noise w_k at time t_k is assumed to be normally distributed, with zero mean, and $(n \times n)$ covariance matrix Q_k . The matrices Φ_k and Q_k are assumed to be known. The $(m \times 1)$ vector z_k is the vector of measurements at time t_k which are used by the filter to estimate x_k . The measurements are linearly related to the states through the $(m \times n)$ measurement matrix H_k .

In addition, the $(m \times 1)$ white additive measurement noise vector v_k corrupts the measurement. It is also assumed to be normally distributed, with zero mean, and having an $(m \times m)$ covariance matrix R_k , assumed to be known. The estimate of state is denoted by \hat{x}_k , with superscript - and + signs denoting the value just before and just after incorporating the measurement obtained at time t_k . The $(n \times n)$ matrix P_k is the covariance of the errors in the estimate of the state (the Kalman filter is said to "generate its own error analysis") and the $(n \times m)$ matrix K_k is the Kalman gain matrix at time t_k , both of which are computed by the filter at each measurement time step t_{k-1}, t_k, \dots . The Kalman gain K_k determines the weighting to be placed on alteration of each component of the estimate of the state vector, given the difference $(z_k - H_k \hat{x}_k^-)$ between the measurement z_k at time t_k and the expected value of that measurement, $(H_k \hat{x}_k^-)$, based on the a priori estimate \hat{x}_k^- of the state.

Define the unknown misalignment vector ϕ to be the (3×1) state vector x_k of the Kalman filter formulation. That is,

$$x_k \equiv \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}_k \equiv \begin{bmatrix} \phi_x \\ \phi_y \\ \phi_z \end{bmatrix}_k \quad (5)$$

When the two ISA's are mounted on a rigid body, the angle ϕ between the case axes is a constant, so the dynamical equation of the Kalman filter formulation is

$$x_k = x_{k-1} \quad (6)$$

in which the state-transition matrix Φ_{k-1} is equal to the identity matrix I . From the physical assumptions stated earlier in regard to rigidity and load stability, there is no process noise in the problem, so $Q_k = 0$.

Let us define the measurement vector z_k of the Kalman filter to be the (3×1) vector obtained by differencing the outputs of the two ISA's at time t_k .

$$z_k \equiv \begin{bmatrix} W_x - W'_x \\ W_y - W'_y \\ W_z - W'_z \end{bmatrix}_k \quad (7)$$

We can use our vector analysis relationship (2) to help us define the observation matrix H_k . Let us work first in vector analysis notation. We drop the k subscript temporarily for notational simplicity.

$$\bar{W} = \bar{i}W_x + \bar{j}W_y + \bar{k}W_z \quad (8)$$

$$\bar{\phi} = \bar{i}\phi_x + \bar{j}\phi_y + \bar{k}\phi_z \quad (9)$$

From (2), and using the determinant for the cross-product,

$$d\bar{W} = \bar{\phi} \times \bar{W} = \begin{vmatrix} \bar{i} & \bar{j} & \bar{k} \\ \phi_x & \phi_y & \phi_z \\ W_x & W_y & W_z \end{vmatrix} \quad (10)$$

$$\bar{i}dW_x + \bar{j}dW_y + \bar{k}dW_z = \bar{i}(\phi_yW_z - \phi_zW_y) + \bar{j}(\phi_zW_x - \phi_xW_z) + \bar{k}(\phi_xW_y - \phi_yW_x) \quad (11)$$

Now convert to matrix notation.

$$\begin{bmatrix} dW_x \\ dW_y \\ dW_z \end{bmatrix} = \begin{bmatrix} \phi_yW_z - \phi_zW_y \\ \phi_zW_x - \phi_xW_z \\ \phi_xW_y - \phi_yW_x \end{bmatrix} \quad (12)$$

Replace the left-hand side using (4) and (7).

$$z = \begin{bmatrix} \phi_yW_z - \phi_zW_y \\ \phi_zW_x - \phi_xW_z \\ \phi_xW_y - \phi_yW_x \end{bmatrix} \quad (13)$$

This is a relationship between z on the left-hand side and the three components of the state vector x on the right. The difficult-looking right side, in which the unknown state vector components seem inextricably scrambled with measured quantities, can be factored into the following simple matrix product:

$$z = \begin{bmatrix} 0 & W_z & -W_y \\ -W_z & 0 & W_x \\ W_y & -W_x & 0 \end{bmatrix} \begin{bmatrix} \phi_x \\ \phi_y \\ \phi_z \end{bmatrix} = \begin{bmatrix} 0 & W_z & -W_y \\ -W_z & 0 & W_x \\ W_y & -W_x & 0 \end{bmatrix} x \quad (14)$$

Comparing (14) with equation (I-2) of Table 1, we are led to define the (3 x 3) coefficient of the x -vector to be the observation matrix H_k :

$$H_k \equiv \begin{bmatrix} 0 & W_z & -W_y \\ -W_z & 0 & W_x \\ W_y & -W_x & 0 \end{bmatrix}_k \quad (15)$$

where the reintroduced subscript k denotes the value of the corresponding quantities at time t_k .

That the state-vector x is linearly related to the measurement z in equation (14) is what makes the Kalman filter both applicable and attractive as the means to solve the alignment problem. To find that the measurement is linearly related to the state is by no means trivial or accidental. If we had defined the state as ϕ and the measurement as W , rather than dW , we would have had a nonlinear problem to deal with, and things would not have worked out nearly so well. The choice of what to call the "state" and what to call the "measurement" is a key aspect of the art in Kalman filter design.

To run the Kalman filter, one needs, besides the measurements, the matrices Φ_k , H_k , Q_k , R_k , an estimate \hat{x}_0 of the initial state-vector x_0 , and an initial value P_0 of the covariance matrix P_k . (In a good filter design, the results after a few measurements will not be sensitive to the initial estimates \hat{x}_0 , generally set to zero, or P_0 , generally set to some fairly large diagonal matrix.) The matrices Φ_k , H_k and Q_k have already been defined. Values for the diagonal elements of R_k are simply twice the variances of the measurement error in each gyro output, the factor of two accounting for the fact that two gyro outputs and hence two errors are introduced into each component of z_k . The "measurement error" in this case is the gyro drift rate, whose variance is denoted by σ_w^2 . All other elements of R_k are zero if it can be assumed that the measurement errors of each gyro are independent of those of the other gyros, usually reasonable. The complete Kalman filter for the solution to this problem is shown in Tables I and II.

OBSERVABILITY

The deterministic observability of the system (that is, the ability to determine the state vector from the given measurement if there is neither measurement noise to corrupt the measurements nor process noise to push the state into uncharted territory) can be tested by the standard observability test, as given, for example, on p. 68 of Ref. 4. When this test is carried out, the results are these: the state vector can be observed given measurements by both ISA's of the angular velocity vector at two measurement times, provided only that the two vectors are not collinear. Thus, in the length of time it takes for the supporting body, say an aircraft, to acquire a roll-rate and then a pitch-rate, a perfect set of instruments could produce data that a Kalman filter could unscramble to obtain the true value of the misalignment. The unequivocal result of the observability test suggests that, even with imperfect gyros, a Kalman filter will provide a good, solid estimate of the state vector in a short time. (Not every Kalman filter designer has the luxury of being able to carry out the observability test, much less

Table I - Summary of discrete Kalman filter equations, general case.

(I-1) System model	$x_k = \Phi_{k-1} x_{k-1} + w_{k-1}, \quad w_k \sim N(0, Q_k)$
(I-2) Measurement model	$z_k = H_k x_k + v_k, \quad v_k \sim N(0, R_k)$
(I-3) Initial conditions	$E[x(0)] = x_0, \quad E[(x(0) - x_0)(x(0) - x_0)^T] = P_0$
(I-4) Other assumptions	$E[w_k v_j^T] = 0$ for all j, k
(I-5) State estimate extrapolation	$\hat{x}_k^- = \Phi_{k-1} \hat{x}_{k-1}^+$
(I-6) Error covariance extrapolation	$P_k^- = \Phi_{k-1} P_{k-1}^+ \Phi_{k-1}^T + Q_{k-1}$
(I-7) State estimate update	$\hat{x}_k^+ = \hat{x}_k^- + K_k [z_k - H_k \hat{x}_k^-]$
(I-8) Error covariance update	$P_k^+ = [I - K_k H_k] P_k^-$
(I-9) Kalman gain matrix	$K_k = P_k^- H_k^T [H_k P_k^- H_k^T + R_k]^{-1}$

Table II - Matrices for Kalman filter. ISA's joined by a rigid body, gyros are used for measurement.

$$\begin{aligned}
 & \text{(II-1)} \quad x_k \equiv \begin{bmatrix} \phi_x \\ \phi_y \\ \phi_z \end{bmatrix}_k \quad \text{(II-2)} \quad H_k \equiv \begin{bmatrix} 0 & W_z & -W_y \\ -W_z & 0 & W_x \\ W_y & -W_x & 0 \end{bmatrix}_k \quad \text{(II-3)} \quad z_k \equiv \begin{bmatrix} W_x & -W_y \\ W_y & -W_x \\ W_z & -W_z \end{bmatrix}_k \quad \text{(II-4)} \quad \Phi_k \equiv I_3
 \end{aligned}$$

$$\begin{aligned}
 & \text{(II-5)} \quad \Delta t_k = t_k - t_{k-1} \\
 & \text{(II-6)} \quad Q_k = 0 \text{ or} \\
 & \text{(II-7)} \quad Q_k = \alpha \Delta t_k I_3 \text{ where} \\
 & \text{(II-8)} \quad \alpha = \text{suitably small constant to introduce fictitious process noise} \\
 & \text{(II-9)} \quad R_k = 2\sigma_w^2 I_3 \text{ where} \\
 & \text{(II-10)} \quad \sigma_w^2 = \text{variance of measurement error at output of each gyro.}
 \end{aligned}$$

$$\begin{array}{lll}
 \text{Dimensions are:} & x_k & 3 \times 1 \\
 & P_k & 3 \times 3 \\
 & H_k & 3 \times 3 \\
 & R_k & 3 \times 1 \\
 & K_k & 3 \times 3 \\
 & z_k & 3 \times 1 \\
 & Q_k & 3 \times 3
 \end{array}$$

Start with initial values \hat{x}_0^+, P_0^+ .
Solve equations (I-5) through (I-9) iteratively, i.e., going through them once at each step with the measurement z_k obtained at that step.

to produce positive results together with the necessary and sufficient conditions for observability to hold.)

If one implemented a filter with $Q_k = 0$ as suggested above, after a few measurements, the filter would be convinced that it had estimated the misalignment angle to a high degree of accuracy, depending on the level of the measurement noise. It would begin to pay less and less attention to further measurements as they came in by reducing the Kalman gain matrix K_k to a very low value. Sometimes, to prevent this, the diagonal elements of Q_k are given a small positive value. This is called "fictitious process noise," and is discussed on pp. 279-280 of Ref. 4. The result is that after many measurements K_k reaches a non-zero, steady-state level and the information in the latest measurement is retained.

NON-RIGID BODY

The analysis is now extended to include the effect of dynamic bending of the body. "Dynamic bending" means that the bending angle changes during the time interval over which the alignment process is carried out. The Kalman filter will be modified to estimate both the static misalignment and the instantaneous dynamic bending angle.

As previously, the body is assumed to undergo angular rotations with respect to inertial space, such as wave- or wind-induced roll, pitch, and yaw, as well as purposeful turning maneuvers. The bending causes an additional angular rotation of the axes of ISA2 relative to ISA1. When this bending changes in time, there is an additional angular velocity measured by the gyros on ISA2 which does not appear at ISA1.

As shown in Fig. 5a, with no bending, the case axes of ISA2 may be misaligned by a small vector angle ϕ relative to the axes of ISA1. When, in addition, the body bends through a small vector angle θ (at the location of ISA2 relative to ISA1), then the total misalignment of ISA2 relative to ISA1 is $(\phi + \theta)$, as in Fig. 5b. While only two dimensions are shown in the figure, with both ϕ and θ being vectors along the Z(z) axis, in the general case ϕ and θ can each have arbitrary components along all three of the axes, X, Y, Z. Furthermore, θ is time-varying. What insight can be brought to bear on the problem before launching into detailed analysis? If there were no bending, then the misalignment could be estimated by the previous method. If there were no misalignment, then the bending could be estimated by simply noting that ISA2 measures a different angular velocity than ISA1, the difference being due to the

fact that ISA2 rotates faster in space than does ISA1 by the rate of change ($\dot{\theta}$) of the bending angle. We could subtract the angular rate of ISA1 from ISA2. The difference would be $\dot{\theta}$. Integration of this angular velocity in a computer would give $\theta(t)$ as a continuous output. (The angle θ is needed to properly subtract the angular rate of ISA2 from ISA1; before this subtraction can be carried out, the two rates must be referred to a common coordinate system. Knowledge of θ permits this to be done, using a feedback loop.)

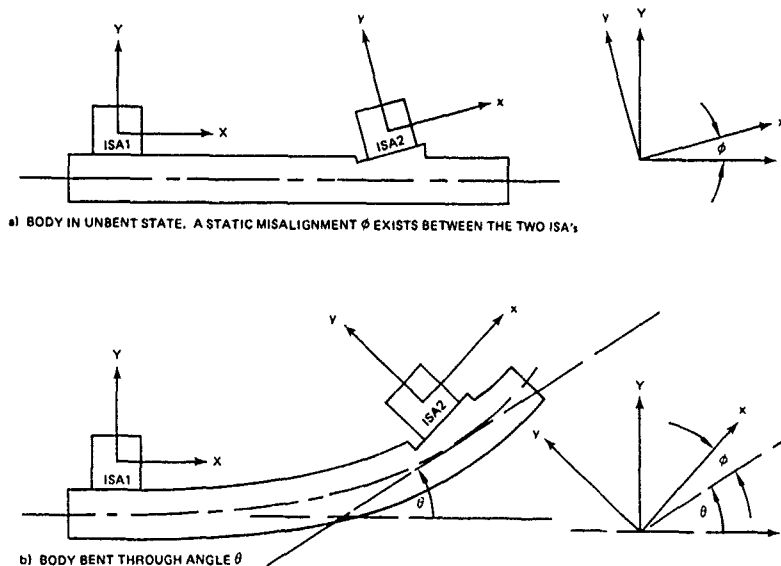


Fig. 5.

Thus, in the absence of $\bar{\phi}$, $\bar{\theta}(t)$ can be found, and vice versa. Since the processes involved are linear, it seems reasonable to expect that when both θ and ϕ are present, their effects can be separated. With that suggestion of pending success, we set out to develop the Kalman filter which will do the job. We shall write the dynamical equations initially as differential equations in continuous variables, and change to difference equations of discrete variables later.

We start by selecting the F-frame components of each of the vectors $\bar{\phi}$ and $\bar{\theta}$ as the states to be estimated by the filter. That is, define x as follows:

$$\begin{array}{ll} x_1 = \phi_x & x_4 = \theta_x \\ x_2 = \phi_y & x_5 = \theta_y \\ x_3 = \phi_z & x_6 = \theta_z \end{array} \quad x = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \\ x_6 \end{bmatrix} \quad (16)$$

In view of the assumption that the static misalignment is constant, the dynamical equations for the first three state variables are, as before,

$$\dot{x}_1 = \dot{x}_2 = \dot{x}_3 = 0 \quad (17)$$

The components of $\bar{\theta}$, as indicated earlier, will be aircraft (or ship) rotation induced by wind (or waves). Hence, they are random variables excited by a random forcing function. Markov processes excited by white noise are of this type. At this point, there is a diversity of models that could be chosen to represent the dynamics of x_4 , x_5 , and x_6 . Thinking of the body connecting ISA1 and ISA2 as being like a mass-spring system, having both inertia and a restoring torque under bending, it seems that we ought to choose at least a second-order noise process to define the motion on each axis. Thus, we are led to consider x_4 , x_5 , and x_6 each to be driven by white noise through a second-order Markov process. We choose the three such processes to be independent, that is, the noise which excites pitch-bending is independent of that which excites roll-bending, etc. It should be emphasized that this is only one model from among many which could be adopted. The model chosen is complex enough to represent the true situation with a fair level of fidelity, yet simple enough to illustrate the development of the Kalman filter. In an actual application, a more accurate bending model, and wind (or wave-) motion model, possibly involving correlation between the three axes, may be desirable, depending upon the accuracy and speed-of-convergence requirements. On the other hand, even the simple second-order Markov processes assumed here may give acceptably good results in many applications.

Having decided to use independent second-order Markov processes for bending in each of the three axes, we find we need to add three more state variables, the derivatives of variables x_4 , x_5 , x_6 . Thus, define

$$\begin{aligned} \dot{x}_7 &= \dot{x}_4 \\ \dot{x}_8 &= \dot{x}_5 \\ \dot{x}_9 &= \dot{x}_6 \end{aligned} \quad (18)$$

Using equation (2.2-82) p. 44, Ref. 4, for the prototype second-order Markov process, the complete set of dynamical equations can now be assembled in the standard Kalman filter form:

$$\begin{aligned} \dot{x}_1 &= 0 \\ \dot{x}_2 &= 0 \\ \dot{x}_3 &= 0 \\ \dot{x}_4 &= x_7 \\ \dot{x}_5 &= x_8 \\ \dot{x}_6 &= x_9 \\ \dot{x}_7 &= -\beta_x^2 x_4 - 2\beta_x x_7 + w_x \\ \dot{x}_8 &= -\beta_y^2 x_5 - 2\beta_y x_8 + w_y \\ \dot{x}_9 &= -\beta_z^2 x_6 - 2\beta_z x_9 + w_z \end{aligned} \quad w \equiv \begin{bmatrix} w_1 \\ w_2 \\ w_3 \\ w_4 \\ w_5 \\ w_6 \\ w_7 \\ w_8 \\ w_9 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ w_x \\ w_y \\ w_z \end{bmatrix} \quad x \equiv \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \\ x_6 \\ x_7 \\ x_8 \\ x_9 \end{bmatrix} \quad (19)$$

where the vector w is a 9×1 vector white noise process with zero mean, normal distribution, and spectral density Q .

$$w \sim N(0, Q) \quad (20)$$

Q is a 9×9 matrix; its 77, 88, and 99 elements are the only non-zero elements in the matrix, and they each represent the spectral density of the white noise w_i that drives the bending of the i th channel ($i = X, Y, Z$). Simple relations exist between the spectral density of the white noise, the constant β_i , and the variance of the output, i.e., the bending angle. Consider, for example, Q_{77} , β_x , and the variance σ_x^2 of $x_4 = \theta_x$. Using the relationships in the table of p. 44, Ref. 4, plus the fact that the ratio of the power spectral density of the output of a linear system to the power spectral density of the input is $G(j\omega)G(-j\omega)$, where $G(j\omega)$ is the system transfer function evaluated at $s = j\omega$, we find that

$$Q_{77} = 4\beta_x^3 \sigma_x^2 \quad (21)$$

Similarly

$$Q_{88} = 4\beta_y^3 \sigma_y^2 \quad (22)$$

$$Q_{99} = 4\beta_z^3 \sigma_z^2 \quad (23)$$

The correlation time τ_i of each random process is simply related to the corresponding β . Again, using the table on p. 44 of Ref. 4,

$$\beta_i = \frac{2.146}{\tau_i} \quad (i = X, Y, Z) \quad (24)$$

We can make an informed guess of the variance of σ_i^2 of the bending angle and the correlation time τ_i on each of the three axes. (In good filter designs, it is frequently found that results are not overly sensitive to the value of τ_i in which case a reasonable guesstimate will suffice.) Then, using equations (21)-(24), we convert these to values for the three non-zero elements of the Q matrix needed to carry out the Kalman filter operations.

Equation (19) can be written in matrix form

$$\dot{x} = Fx + w \quad (25)$$

where F is the 9×9 matrix of constants:

$$F = \begin{bmatrix} 0_3 & & & & & & & & \\ & 0_3 & & & & & & & \\ & & 0_3 & & & & & & \\ & & & 0_3 & & & & & \\ & & & & -\beta_x^2 & 0 & 0 & -2\beta_x & 0 \\ & & & & 0 & -\beta_y^2 & 0 & 0 & -2\beta_y \\ & & & & 0 & 0 & -\beta_z^2 & 0 & 0 \\ & & & & & & & 0 & -2\beta_z \end{bmatrix} \quad (26)$$

and where 0_3 and I_3 are the (3×3) null and identity matrices, respectively.

Our physical reasoning has led us to anticipate that the key in the measurement process is the difference between the angular velocity of the two ISA's. This suggests that, as previously, we define the measurement to be $d\bar{W}$. In vector form, this is:

$$d\bar{W} = \bar{W} - \bar{W}' = \bar{i}[W_x - W_x] + \bar{j}[W_y - W_y] + \bar{k}[W_z - W_z] \quad (27)$$

With matrices, the measurement takes the form

$$z \equiv \begin{bmatrix} z_1 \\ z_2 \\ z_3 \end{bmatrix} \equiv \begin{bmatrix} W_x - W_x \\ W_y - W_y \\ W_z - W_z \end{bmatrix} \quad (28)$$

It is necessary to know how a perfect (i.e., noise-free) measurement would be related to the states. If it can be shown to be linearly related, then we will have precisely what is needed to implement a Kalman filter. When this relationship is formulated, it is found, indeed, to be linear! It is given by

$$z = \begin{bmatrix} 0 & W_z & -W_y \\ -W_z & 0 & W_x \\ W_y & -W_x & 0 \end{bmatrix} \begin{bmatrix} \phi_x + \theta_x \\ \phi_y + \theta_y \\ \phi_z + \theta_z \end{bmatrix} - \begin{bmatrix} \dot{\theta}_x \\ \dot{\theta}_y \\ \dot{\theta}_z \end{bmatrix} \quad (29)$$

or in terms of state variables,

$$\begin{bmatrix} z_1 \\ z_2 \\ z_3 \end{bmatrix} = \begin{bmatrix} 0 & W_z & -W_y \\ -W_z & 0 & W_x \\ W_y & -W_x & 0 \end{bmatrix} \begin{bmatrix} x_1 + x_4 \\ x_2 + x_5 \\ x_3 + x_6 \end{bmatrix} - \begin{bmatrix} x_7 \\ x_8 \\ x_9 \end{bmatrix} \quad (30)$$

Alternatively, this can be written as follows to put the H-matrix into full view, where the effect of imperfect measurements is now appended as additive white noise.

$$\begin{bmatrix} z_1 \\ z_2 \\ z_3 \end{bmatrix} = \begin{bmatrix} 0 & W_z & -W_y & 0 & W_z & -W_y & -1 & 0 & 0 \\ -W_z & 0 & W_x & -W_z & 0 & W_x & 0 & -1 & 0 \\ W_y & -W_x & 0 & W_y & -W_x & 0 & 0 & 0 & -1 \end{bmatrix} x + \begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix} \quad (31)$$

The coefficient of x is the (3×9) observation matrix H . The R -matrix dimensions are still (3×3) , since the z -vector as now defined still has 3 components. As before, if the gyros are identical and are presumed to have statistically identical noise outputs, then the three diagonal elements of R are equal; the off-diagonal elements are zero if the errors of each gyro are independent of the errors of all of the others.

Equations (29) and (31) are the matrix equivalent of the following vector relationship:

$$d\bar{W} = (\bar{\phi} + \bar{\theta}) \times \bar{W} - \dot{\bar{\theta}} \quad (32)$$

where the first term on the right side represents the cross-product effect for a total instantaneous misalignment $(\bar{\phi} + \bar{\theta})$ between the two ISA's, corresponding to equation (10) above for the rigid-body situation. (Also, compare the corresponding matrix representations, eqs. (29) and (14).) The second term is the correction for the instantaneous angular velocity of ISA2 relative to ISA1. If such an angular velocity exists, the outputs of ISA1 and ISA2 would be expected to differ even if $\bar{\phi} = \bar{\theta} = 0$. Since 1) instantaneous misalignment $(\bar{\phi} + \bar{\theta})$, and 2) angular bending-rate $(\dot{\bar{\theta}})$, affect $d\bar{W}$ in a linear fashion, the effect due to the simultaneous presence of both is the superposition of their individual effects, hence (32).

All of the parts for a Kalman filter design are present: F is given in (26), Q is defined in (21)-(23), z is defined in (28), and H is defined by (31). In practice, one would convert from the continuous to the discrete Kalman filter, taking a sequence of angular velocity (or incremental angle) measurements every Δt seconds. The conversion from continuous to discrete form is straightforward. Ref. 4, p. 77 gives the following approximate equations which can be used to carry out the discretization:

$$\Phi_k = I + F \Delta t_k \quad (33)$$

$$Q_k = Q \Delta t_k \quad (34)$$

Equation (33) is the first-order approximation to the state-transition matrix Φ_k , and is valid when Δt_k is much smaller than the shortest period in the natural modes of the system described by $\dot{x} = Fx$. This is surely the case in the present application. R_k for this filter is the same as in the rigid-body case.

This completes the Kalman filter design for the case in which the body supporting the two strap-down packages is assumed to bend, and the bending angle changes over the time period in which the alignment process is carried out. This problem, too, fits perfectly into the Kalman filter mold. When "state" and "measurement" are defined in the ways shown above, both the dynamic and measurement processes are linear in the state. Hence, we fully expect that a Kalman filter will successfully solve the problem; it will be able to obtain an estimate of the current value of the total angular misalignment vector (the sum of static misalignment plus dynamic bending) between two ISA's. What should happen is that the filter estimate of $\bar{\phi}$ will converge to essentially the true value of $\bar{\phi}$, after which this estimate will not change very much. From then on, the main activity of the filter will be to keep up with the continually changing $\bar{\theta}(t)$. For the reason described previously, it may be useful to put a small amount of fictitious process noise into states $i = 1, 2, 3$ by setting $Q_{k,ii}$ equal to some small positive number.

With the measurement being linear combinations of the states, and with high observability afforded by base-motion angular velocity inputs along different directions (as described earlier), one can expect that the filter: 1) will converge rapidly, 2) will be relatively insensitive to noise parameters of the model (such as the correlation times τ_i), and 3) will be relatively insensitive to mismatch between the dynamical model and the real world (such as the dynamics of bending and the gyro noise dynamics). The last point, if indeed true, implies that the detailed form of equation selected to model the gyro measurement errors, bending dynamics, etc. will not have a first-order effect on the performance of the filter. These conjectures should be verified by simulation for any given application. A nine-state model is parsimonious; a six-state model would be the absolute minimum for estimating three components each of $\bar{\phi}$ and $\bar{\theta}$. Yet, the nine-state model is probably large enough to obtain acceptable accuracy in

many applications, for the reasons enumerated above. At the same time, a nine-state Kalman filter is small enough to be readily implemented in today's state-of-the-art of computing technology.

Tables I and III give the complete discrete Kalman filter for estimating the fixed misalignment and variable bending using gyro outputs of two ISA's. It should be mentioned again that both Table II and III describe the discrete Kalman filter for the situation when the output of a gyro is a sample of angular velocity at $t = t_k$. Slight modifications are required to handle the case when the output is an increment of angle.

ESTIMATING GYRO DRIFTS

The filter just described can be augmented to estimate the drift rate of the gyros about the three components axes of each ISA. To do this, one defines six new state-vector components as follows:

$$\begin{aligned} x_{10} &= \dot{\eta}_x \\ x_{11} &= \dot{\eta}_y \\ x_{12} &= \dot{\eta}_z \\ x_{13} &= \dot{\eta}_x \\ x_{14} &= \dot{\eta}_y \\ x_{15} &= \dot{\eta}_z \end{aligned} \quad (35)$$

where η is the drift rate along the subscripted axis. A dynamic model must be adopted for these states. One possibility is to assume that the drift rates are constant, in which case the dynamics are

$$\dot{x}_i = 0 \quad i = 10, \dots, 15 \quad (36)$$

As another alternative, one could assume that the drift rates change slowly in time due to white noise inputs. Such a model would be

$$\dot{x}_i = 0 + w_i \quad i = 10, \dots, 15 \quad w_i \sim N(0, q_i) \quad (37)$$

This leads to a random walk behavior for these state variables. A still more elaborate model for the gyro drift rate states would be to model each one as a first-order Markov process

$$\dot{x}_i = -\beta_i x_i + w_i \quad w_i \sim N(0, q_i) \quad i = 10, \dots, 15 \quad (38)$$

This has the advantage that the x_i 's tend toward zero in the absence of the white noise w_i input, rather than the unbounded random walk behavior of (37). This model can be accommodated with a 15-state filter, using the states which have been defined. Still more elaborate drift-rate models are possible, but these will require a larger state-vector. If all gyros had identical designs, there would be no reason to suspect that the statistical behavior of any one would differ from that of any other, so identical β_i 's and q_i 's would be chosen for the six channels. The q_i 's become diagonal elements of Q . The corresponding discrete matrix Q_k is then found from (34).

Under what conditions would it be worth while to attempt to estimate the gyro drift rates? The answer to this question will be stated in terms of extremes. If the drift rate of the gyros is low, if the time allowed for the alignment process is short (say, 5 minutes or less), if the body experiences large angular velocities during the alignment time, and if the desired accuracy of the alignment procedure is modest, then it probably is not worthwhile to attempt to estimate the gyro drifts; hence states 10-15 can be omitted. On the other hand, if the gyro drift rate is high (particularly of the second ISA compared to the first), if the time allowed is long (say, an hour or more), if the angular velocity input is low, and if the ultimate in alignment accuracy is desired, then the gyro drift states ought to be included. In between these extremes, one would have to specify numerical values and perform a detailed simulation in order to determine the relative improvement in accuracy obtainable by including the additional states.

ACCELEROMETERS, RIGID BODY

If the ISA's have accelerometers along all three axes, as well as gyros, this affords an opportunity for more measurements, as well as introducing more errors. The measurements made by the three accelerometers of one ISA are the three components along case axes of the specific force vector \bar{f} , defined as the non-gravitational force per unit mass acting on the case. Craft operating close to the earth (and not in free-fall) experience the gravity field of the earth. The case must be supported to prevent its falling. The ISA measures this support specific force, even in the absence of maneuvers. This specific force \bar{f} can be one of the two non-collinear vectors measured by the ISA's to achieve the observability of the filter. The other vector can be another value of \bar{f} at a later time (and in a different direction, due, say, to vehicle acceleration) or it could be an angular velocity measured either simultaneously or later or earlier than \bar{f} .

To process accelerometer measurements, we define a different measurement vector z_a and a different observation matrix H_a , where the subscript a denotes "accelerometer." For the rigid-body case, the measurement z_a would be the difference in specific force measurements of the two ISA's, measured at the same time.

$$z_{a,k} = \begin{bmatrix} df_x \\ df_y \\ df_z \end{bmatrix}_k = \begin{bmatrix} f_x^x \\ f_y^x \\ f_z^x \end{bmatrix}_k - \begin{bmatrix} f_x^y \\ f_y^y \\ f_z^y \end{bmatrix}_k \quad (39)$$

(As in the case of the gyros, in practice, we may prefer to define the measurement at t_k as $\bar{f} \Delta t_k$, where Δt_k is the time interval since the last measurement, and the product $\bar{f} \Delta t_k$ is the output of integrating accelerometers at $t = t_k$. Slight changes in the equations are then required.) When using accelerometer

measurements, and assuming that the accelerometer axes are coincident with the gyro axes in each ISA (for present purposes we assume that it is only the entire case of ISA2 which might be misaligned) then the observation matrix H_a is the coefficient of x in

$$z_{a,k} = \begin{bmatrix} 0 & f_z & -f_y \\ -f_z & 0 & f_x \\ f_y & -f_x & 0 \end{bmatrix}_k x_k = H_{a,k} x_k \quad (40)$$

which is the matrix equivalent of the vector cross-product,

$$d\vec{f} = \vec{\phi} \times \vec{f} \quad (41)$$

the accelerometer counterpart of (2).

Thus, one could implement the Kalman filter given in Table II, having the same state variables, but amending it to include the additional observation matrix $H_{a,k}$ and measurement set $z_{a,k}$ in (40) and (41). One would process first the angular velocity measurement z_k and then the accelerometer measurements $z_{a,k}$ at time t_k , using the appropriate observation matrix for each.

In addition, a correction term may have to be added to (40) to account for difference in acceleration at ISA2 relative to ISA1. For example, if one ISA is at the center of gravity of the body, and the other is removed from this point by some distance, then the second ISA will experience a centripetal acceleration not experienced by the first whenever the craft rotates.

ACCELEROMETERS, NON-RIGID BODY

In this case also one could use the (9 x 1) set of state vectors defined by (16) and (18) for the companion gyro-only case, but adding a new measurement $z_{a,k}$ as defined by (39) and a new observation matrix $H_{a,k}$ defined by (42),

$$z_{a,k} = \begin{bmatrix} 0 & f_z & -f_y & 0 & f_z & -f_y & 0 & 0 & 0 \\ -f_z & 0 & f_x & -f_z & 0 & f_x & 0 & 0 & 0 \\ f_y & -f_x & 0 & f_y & -f_x & 0 & 0 & 0 & 0 \end{bmatrix}_k x_k = H_{a,k} x_k \quad (42)$$

except that now one would also have to add a correction which accounts for the difference in the accelerations at the two locations caused by bending angle $\bar{\theta}$ and bending rate $\dot{\theta}$ as well as angular velocity and angular acceleration of the body.

ESTIMATING ACCELEROMETER BIASES

The addition of accelerometers introduces new sources of error as well as new measurements. Thus, one may wish to add a new state variable for each accelerometer to represent its bias. Then the filter will estimate these biases along with all else. These could be modelled as constants, random walks, or first-order Markov processes, like the gyro drifts. One would have a 21-state filter if one wanted to estimate gyro drifts and accelerometer biases as well as static misalignment and dynamic bending, assuming first-order Markov processes for the drifts and biases.

Other error parameters of the ISA's can be estimated by adding more states to the filter in similar fashion. However, as the number of states of the filter increases, the computational load gives up at a higher power, observability may suffer, and the filter may become increasingly sensitive to the accuracy of the model.

My own preference would be to try to solve any given design problem and mission requirement with the 9-state filter of Table III, without estimating either gyro drifts or accelerometer biases. If that design met system requirements, there would be no further need to complicate the filter.

ACKNOWLEDGEMENT

This problem was suggested to the author by Donald W. Doherty and Donald H. Lackowski of the Naval Ocean Systems Center, San Diego. The author wishes to acknowledge their encouragement.

REFERENCES

1. Carta, D.G. and Lackowski, D.H., "Estimation of Orthogonal Transformations in Strapdown Inertial Systems," IEEE Transactions on Automatic Control, Vol. AC-17, No. 1, Feb., 1972, pp. 97-100.
2. Browne, B.H. and Lackowski, D.H., "Estimation of Dynamic Alignment Errors in Shipboard Firecontrol Systems," Proceedings of IEEE Conference on Decision and Control, Dec., 1976.
3. Yamamoto, G.H. and Brown, J.T., "Design, Simulation and Evaluation of the Kalman Filter used to Align the SRAM Missile," AIAA Paper No. 71-948, AIAA Guidance, Control, and Flight Mechanics Conference, Aug. 16-18, 1971.
4. Gelb, A. (ed.). Applied Optimal Estimation, M.I.T. Press, Cambridge, Mass., 1974.

Table III - Matrices for Kalman filter. ISA's joined by a non-rigid body; gyros are used for measurement

$$(III-1) \quad x_k^T \equiv [\phi_x \phi_y \phi_z \theta_x \theta_y \theta_z \dot{\theta}_x \dot{\theta}_y \dot{\theta}_z]_k$$

$$(III-2) \quad z_k \equiv \begin{bmatrix} W_x - W_x \\ W_y - W_y \\ W_z - W_z \end{bmatrix}_k$$

$$(III-3) \quad H_k \equiv \begin{bmatrix} 0 & W_z & -W_y & 0 & W_z & -W_y & 0 & 0 & 0 \\ -W_z & 0 & W_x & -W_z & 0 & W_x & 0 & 0 & 0 \\ W_y & -W_x & 0 & W_y & -W_x & 0 & 0 & 0 & 0 \end{bmatrix}_k$$

$$(III-4) \quad \Phi_k \equiv \begin{bmatrix} I_3 & 0_3 & 0_3 & 0_3 & 0_3 & 0_3 \\ 0_3 & I_3 & 0_3 & 0_3 & 0_3 & 0_3 \\ -\beta_x^2 \Delta t_k & 0 & 0 & 1-2\beta_x \Delta t_k & 0 & 0 \\ 0_3 & 0 & -\beta_y^2 \Delta t_k & 0 & 0 & 1-2\beta_y \Delta t_k \\ 0 & 0 & 0 & -\beta_z^2 \Delta t_k & 0 & 0 & 1-2\beta_z \Delta t_k \end{bmatrix}$$

$$(III-5) \quad Q_k \equiv \begin{bmatrix} a \Delta t_3 I_3 & 0_3 & 0_3 & 0_3 & 0_3 & 0_3 \\ 0_3 & 0_3 & 0_3 & 0_3 & 0_3 & 0_3 \\ 0_3 & 0_3 & 4\beta_x^3 \sigma_x^2 \Delta t_k & 0 & 0 & 0 \\ 0 & 0 & 0 & 4\beta_y^3 \sigma_y^2 \Delta t_k & 0 & 0 \\ 0 & 0 & 0 & 0 & 4\beta_z^3 \sigma_z^2 \Delta t_k & 0 \end{bmatrix}$$

$$(III-6) \quad \sigma_i^2 \equiv \text{variance of bending angle on axis } i$$

$$(III-7) \quad \tau_i \equiv \text{correlation time of bending on axis } i$$

$$(III-8) \quad \beta_i = 2.146/\tau_i$$

$$(III-9) \quad R_k = 2\sigma_w^2 I_3$$

$$\left. \begin{array}{l} (III-6) \\ (III-7) \\ (III-8) \end{array} \right\} i = X, Y, Z$$

COMBINING LORAN AND GPS -- THE BEST OF BOTH WORLDS

by

Paul Braisted, Ralph Eschenback and Anil Tiwari
Trimble Navigation
585 North Mary Ave.
Sunnyvale, CA 94088
United States

ABSTRACT

Loran and GPS are both well known navigation systems, and both have advantages and disadvantages. By combining both, the navigator can get the advantages of both and minimize the disadvantages of each.

When GPS becomes fully operational, it will replace virtually all other navigation systems, but until that time (late 1990), there will be substantial lapses of coverage. It is during this time that Loran provides the backup. This combination is the first system that can provide the transoceanic vessel with real time, accurate navigation information at both ports and periodic accurate navigation information en-route.

This paper will review the two systems and then show how the whole is greater than the sum of the parts. In particular, data will be presented to show how the combined product can yield better navigation information than either system by itself.

SYSTEM DESCRIPTIONS

Complete system descriptions exist elsewhere in the literature (References 1, 2, 3, 4) so only a pertinent overview will be given here. This overview will only allow the reader to see the merits of combining the two systems.

LORAN

Loran-C (Long Range Navigation) was developed during the late 1950's and early 1960's by the U. S. Department of Defense. The system is comprised of about 15 chains or GRI's (Group Repetition Interval) which cover about 70% of the coastlines in the northern hemisphere of North America. A single chain will cover between one and two thousand miles of coastline to a range of about one thousand miles from the coast. Even with this coverage, there are still large gaps in the mid-continent and mid-oceanic regions.

A single chain is made up of a master and two or more slaves. A receiver computes its position by measuring the relative time of arrival of the signal from the master and one slave. This time difference (TD) places the user on one Line-Of-Position (LOP). A measurement from another slave gives two TD's and the intersection determines the user's position.

This all sounds good, but there are several serious problems which make the system less than 100% reliable. These are discussed in the following sections.

ASF - Additional Secondary phase Factor is an error correction term which is applied to the raw TD to correct for the reduced propagation velocity of the Loran signals over land. This term varies as a function of the conductivity of the path over which the signal passes. Most receivers model this conductivity and attempt to correct for this error, but correction below .2nm is rarely achieved. Also this term varies seasonally and this seasonal effect is rarely modeled.

Cycle Slip - No receiver is immune to the difficulties involved with selecting the correct cycle from which to do the time measurement. Under most conditions there is no problem selecting the correct 3rd cycle, but when the signal is weak this can be a very difficult task. When a cycle slip occurs, a position error will result with a magnitude of up to 5 miles. Most receivers allow the user to force the cycle choice to a different cycle. This forces the receiver to choose a particular cycle, but the user must know where he is in order to make this choice.

Ambiguity - An ambiguity occurs whenever the two LOP's cross in two locations. When this occurs, the receiver tries a third LOP (if it is available) to resolve the ambiguity. If this is not available, the user must intervene and specify which is the correct solution. Once again, the user must know where he is in order to get the Loran tracking correctly.

GRI - Although not a serious problem, the user must know the GRI for his area in order to have the Loran work at all. In an airborne application, GRI's can change rapidly, and the user must continuously be aware of these changes.

In spite of all of these problems, Loran is a very useful and effective system. It would be nice, however, if these problems could be reduced or even eliminated. As we will see, GPS can be used to minimize these problems.

GPS

The GPS system was developed by the US Department of Defense starting around 1973. This is a satellite based system where a user measures the pseudorange from four satellites and then computes his position and time. When the system is fully operational, in late 1990, it will give better than 50 meter absolute accuracy and RMS errors of less than 10 meters. At present, however, this level of accuracy is only available about five hours a day. This five hour period moves earlier every day by four minutes.

By assuming that the user is on the surface of the earth, only three satellites need to be used to determine position. This extends the usable time from four to six hours as seen in Figure 1. This two dimensional (2D) coverage will not be continuous until late 1989, and until then the GPS system alone will not be adequate for the practicing navigator.

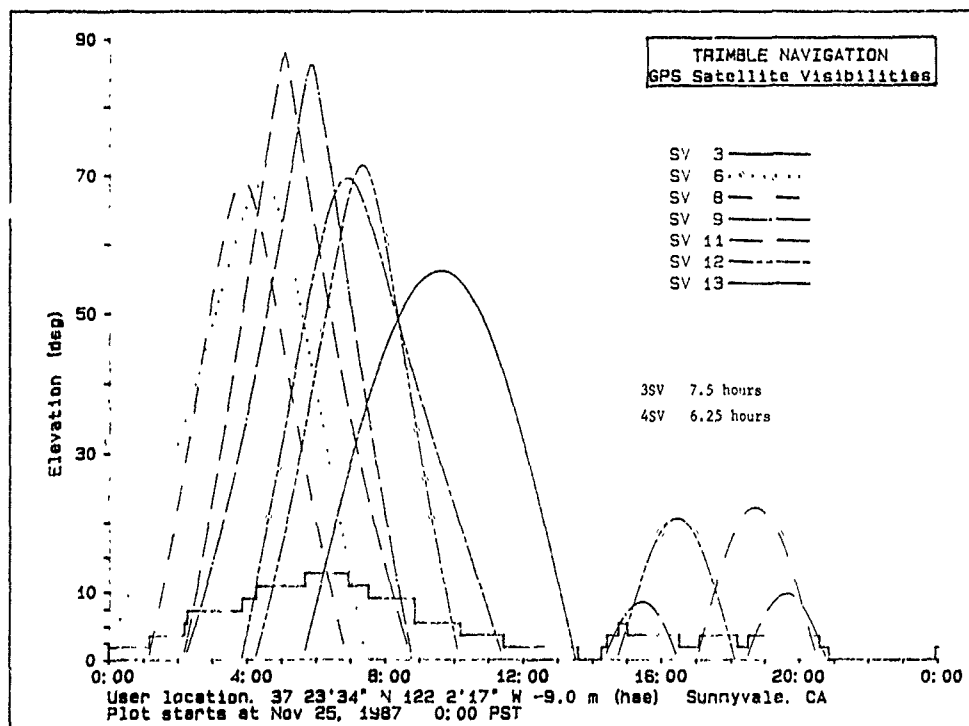


Figure 1. Typical GPS Coverage

The signal transmitted by the satellites is at 1.575 GHz, and is much easier to demodulate than the analog zero crossing of Loran, but a problem still exists. This is called the integer millisecond problem. Since the code is one millisecond long, it cannot determine which millisecond we are tracking with respect to the satellite. To solve this problem, the user must supply an approximate starting position to the GPS receiver. This starting position must be accurate to about three degrees in latitude and longitude.

SYMBIOSIS

This section looks at how the two systems can assist with the deficiencies of the other. Let's first look at how the GPS can aid the Loran.

GPS Aids LORAN

ASF Corrections - As mentioned before, the Loran signal propagation velocities vary as a function of surface conductivity. The models used in most receivers attempt to correct for these errors, but it is impossible to do this very accurately because of the daily and seasonal variation in this conductivity. By using the GPS as an accurate reference, the mean correction for a given day can be computed and used to correct the Loran fixes when the GPS system is not available.

As shown in the typical data in Figure 2, the mean position error between the two systems is about 200 meters, but the RMS of each of the Loran is about 20 meters and the RMS of the GPS is about 5 meters. The Loran data was taken every half hour for a 24 hour period.

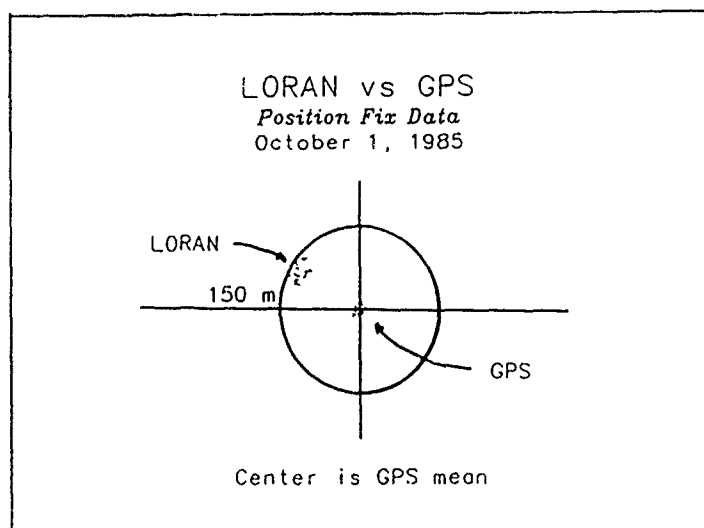


Figure 2. Typical Loran and GPS Position Fixes

Cycle Slip - Due to many causes, a Loran Receiver may have a difficult time determining the correct cycle of the Loran pulse. One of the primary causes is a very weak signal. This can be caused by signal blockage or extreme range to the transmitter. If GPS signals are available, the correct cycle can be determined very accurately, and the Loran can be trusted to give correct answers even when GPS subsequently disappears.

Ambiguity - Once again, GPS can be used to eliminate an ambiguity caused by only being able to receive two slave stations. As with cycle slip, once the ambiguity has been resolved, the Loran can be trusted to give correct answers.

Loran Aids GPS

Initialization Problem - from a cold start, GPS needs to have an approximate location (within 3 degrees latitude and longitude) in order to determine the correct position. If the user is in a Loran coverage area, then the initialization position can be provided by the Loran.

Systems Aid Each Other

There are two ways that the systems can aid each other. The first has to do with the fact that one system completely covers time and the other covers space. That is to say, the GPS gives global coverage but not continuous, and the Loran gives continuous coverage, but not global. Thus the two systems compliment each other in this regard.

The second way that they aid each other concerns what we call the extended coverage capability. With some modifications in software, a position fix can be obtained by taking only two satellites and two Loran transmitters. Thus with one TD and two satellites a position fix can be obtained. This solution extends the coverage area of Loran and the coverage time of GPS.

OTHER COMBINATIONS

GPS/INS - This combination has several attractive features. If the GPS is not available, the INS can provide a solution, and with GPS, the INS can be continuously calibrated. This is quite attractive in a jamming environment since the INS does not require any external signals. For a marine application, however, this advantage is not pertinent. In addition, this solution is quite expensive as even low cost INS's cost more than \$25,000.

GPS/Omega - This solution is also expensive. Even though Omega is global, it does not have the accuracy that Loran does, and GPS provides the global coverage in any case. Thus this combination is both more expensive and less accurate than the LORAN/GPS.

GPS/Transit - The problem with this combination is that the capabilities of both systems overlap. Both systems provide global coverage and both have time gaps in their coverage. What transit needs most is accurate velocity information, and when GPS could provide this, a GPS solution is available, and thus a transit fix is not necessary.

TYPICAL SCENARIOS

Coastal User - For the typical coastal user, the LORAN/GPS combination has many benefits. When GPS is available, much more accuracy is achievable. Since the Loran ASF corrections will be calibrated by GPS, the Loran will be more accurate when the GPS is not available. Lastly, because of the aiding from the GPS, the Loran will have greatly reduced errors caused by cycle slips and ambiguities.

Oceanic User - The oceanic user gets all of the benefits of the coastal set when near port at both ends of the journey. In the middle of the trip, when Loran is not available, GPS can provide accurate periodic navigation. With GPS, set and drift can be computed in the open ocean, and more direct headings and routings can thus be achieved. This is not possible with a system that provides only periodic fixes. Real time navigation as provided by GPS is necessary for this computation. Lastly, the oceanic user can get the extended range capability available only to the user of the LORAN/GPS combination.

SUMMARY

In summary, we have looked at the two systems, GPS and LORAN, and found that both have some deficiencies especially in the interim period before all the GPS satellites are launched. More importantly, however, the advantages of one system can compensate for the disadvantages of the other giving the navigator a more reliable and more accurate system than either system alone can provide. In particular, Loran helps with initialization of GPS and gives continuous coastal coverage. GPS aids Loran by assisting with cycle selection and ambiguities. It also improves Loran accuracies by calibrating ASF factors. In addition it gives at least 6 hours per day of global coverage.

We have also looked at other combinations and shown them not to be as effective for the marine user as the LORAN/GPS. Until GPS becomes fully operational in late 1990, the best solution appears to be a LORAN/GPS.

REFERENCES

1. Institute of Navigation Special Edition on the Global Position System, papers published in "Navigation" (Journal of the Institute of Navigation), 1980.
2. Institute of Navigation Special Edition on the Global Position System, papers published in "Navigation" (Journal of the Institute of Navigation), Vol. 2, 1984.
3. Loran-C Users Handbook, Department of Transportation, Coast Guard, Comdtinst M16562.3, 1980.
4. Eschenbach, R.F., "GPS, the Coming Revolution in Marine Navigation", NMEA News, Vol. 12, No. 1, Jan-Feb. 1985.

THE INTEGRATION OF MULTIPLE AVIONIC SENSORS AND TECHNOLOGIES FOR FUTURE MILITARY HELICOPTERS

BY
ALBERT J. SHAPIRO
VICE PRESIDENT, ELECTRONIC PROGRAMS

ELECTRONIC SYSTEMS DIVISION
THE SINGER COMPANY
164 TOTOWA ROAD
WAYNE, NEW JERSEY
07474-0975
UNITED STATES

SUMMARY

The expanding role of the helicopter in the battlefield environment has burdened the pilot with missions of greater complexity and risk with a concomitant increase in pilot workload. Navigation of the helicopter is an essential supportive element to the prime mission and has been until recent years, a significant contributor to the workload. Technological advances in navigational electronics such as Doppler navigation radar, computers, integrated avionic control and display systems, etc., now can provide automated navigation with vital benefits in cost, size, weight and power, which permit incorporation of these advances into the helicopter. Cost reductions are particularly important since helicopters are used in large quantities in modern military forces. Multi-sensor navigation systems already available and in use in helicopters are discussed, followed by a review of the system trade-offs and considerations leading to new systems that use more advanced digital electronic techniques to achieve the goals of reduced pilot workload and improved performance with minimum size, weight, and cost. The beneficial impact of ongoing technological advances in improving the operating capabilities of future avionics systems is indicated.

IMPACT OF HELICOPTER MISSION REQUIREMENTS ON NAVIGATION

The military helicopter has been given a continuously expanding role both on land and at sea. Helicopter missions include rescue, reconnaissance, troop delivery, attack and weapon delivery. These missions are often performed under active battlefield stress and may require the pilot's full attention for the accomplishment of the mission objective. To permit the pilot to cope with the workload imposed by more demanding mission tasks, it has become necessary to relieve him, through automation, of some of those supporting functions that have been historically performed manually. One of those supporting functions, which technology has permitted to be automated, is precise navigation.

Under favorable conditions of visibility and unrestricted altitude, the pilot workload associated with manual map and compass navigation may be tolerable. However, the battlefield environment no longer permits the helicopter to fly at the 300 to 500-foot altitudes that allow the pilot or navigator to navigate using simple visual terrain and map correlation. Experience in recent conflicts has proven the vulnerability of the helicopter when it is exposed at these altitudes. Ultra-low altitude flight (under 50 feet) decreases the vulnerability of this craft to enemy ground fire, but makes VFR flying by use of map and compass difficult. Any tactical incident will almost surely cause the pilot or navigator (if there is one) to lose track of his position and become disoriented, even in clear weather. This, coupled with the expectation that battlefields of the future will be active at night and in inclement weather, makes map and compass dead reckoning navigation highly impractical.

Figure 1(a) depicts the functions which, in the past, had to be performed manually by a pilot in order to navigate. He served as a data gatherer and computer to combine the output of several sensors to derive the navigation output. Air speed, attitude, heading, drift angle, and altitude are the major inputs that he must read and combine to plot his position on a map. He eliminated accumulated errors in his dead reckoning computation by periodic updates provided by visual identification of local terrain features, a procedure that is highly ineffectual at night, in poor visibility, or at low altitudes. Manual dead reckoning is an acceptable technique when missions and helicopters are simple, but with the increasing workload, stress, and confusion during today's tactical situations, it no longer is adequate.

These problems demanded the development of integrated avionics to provide automatic navigation to reduce pilot workload and supply him with position, steering and guidance information and true ground velocity for many purposes besides navigation. Considerations of survivability and the possibility of providing these outputs using onboard sensors favored a self-contained navigation system as opposed to ground-based aids which could be jammed, spoofed or subject to attack.

DEVELOPMENT OF CURRENT RESPONSE TO BASIC NAVIGATION NEEDS

AN INTEGRATED, AUTOMATIC HELICOPTER NAVIGATION SYSTEM

The system considerations discussed above indicate the need for an automatic navigation system that would, in effect, replace the pilot's manual navigation task with a com-

puter. This allows one to take advantage of the computer's greater capability by adding additional navigation sensors that are more accurate and that further reduce his workload. For example, it now becomes possible to replace the airspeed sensor, which requires the pilot to estimate wind magnitude and direction, with a direct, automatic groundspeed measuring sensor such as a Doppler radar. Figure 1(b) shows the result of applying these new capabilities to the navigational procedures outlined in Figure 1(a).

The navigation and guidance data generated by the computer shown in Figure 1(b) must be displayed in a manner that minimizes the amount of interpretation required by the pilot to translate them into helicopter control actions. The most desirable approach is to enable the pilot to enter the coordinates of destinations or targets prior to or during flight, and to cause the computer to generate a left-right steering signal which, when zeroed, causes the helicopter to fly toward the selected destination. The approach greatly reduces the time spent by the pilot with a hand-held map, and the greater accuracy of such a navigation system reduces the required frequency of position updates and hence, further reduces the pilot's map-reading tasks.

The U.S. Army, recognizing the need for an automatic navigation system such as that shown in Figure 1(b), chose a Doppler Radar Velocity Sensor (DRVS) for the direct, automatic groundspeed measuring sensor. The selection of this type of a velocity sensor is appropriate for several reasons:

- It is self-contained.
- It can provide accurate data throughout the speed/altitude profile of a helicopter, namely, from negative (backward) flight through hover to positive (forward) flight.
- Accuracy is independent of time and hence, of the length of the mission.
- It provides instantaneous reaction time.
- It is all solid state and has no moving parts.
- A Doppler radar has a lower Life-Cycle Cost (LCC) in comparison to other types of velocity sensors, and thus, is affordable for large quantities of helicopters.

Singer had already begun development of a Doppler system having these characteristics when the Army translated these requirements into a specification. The development was completed under Army sponsorship to produce the AN/ASN-128 Lightweight Doppler Navigation System shown in Figures 2 and 3.

The Receiver-Transmitter Antenna (RTA) and Signal Data Converter (SDC) form the DRVS. The RTA contains the antenna with its integral radome, RF components for transmitting and receiving electromagnetic signals, and electronics to condition the resultant signals. The SDC receives the conditioned signals from the RTA, measures the Doppler shifts, and converts these into the digital form required by the computer. The Computer Display Unit (CDU) contains a custom LSI digital computer and memory that perform all navigation and guidance computations, and displays the results on its front panel. Entry of destination coordinates and other data are made via a keyboard, while control switches are provided for mode and display control selection. The Steering-Hover Indicator Unit (SHIU) displays guidance data to the pilot in "analog" form, that is, as pointers and needles, although it also contains a numeric readout of distance-to-go. The SHIU is an optional addition to the system.

The AN/ASN-128 provides accurate navigation over the flight parameters listed below:

Along-heading velocity	-	-50 to +350 knots (-93 to +650 km/h)
Cross-heading velocity	-	± 100 knots (± 185 km/h)
Vertical velocity	-	$\pm 5,000$ ft/min ($\pm 1,500$ m/min)
Altitude	-	0 to 10,000 ft (0 to 3,000 m) above ground level
		<u>Land</u> <u>Water</u>
Attitude - Pitch	-	$\pm 30^\circ$ $\pm 20^\circ$
Roll	-	$\pm 45^\circ$ $\pm 30^\circ$

Accuracy of the AN/ASN-128 can be stated for the DRVS, alone, and for the computed position after Doppler velocity data are combined with heading from the on-board heading reference.

Doppler Radar Velocity Sensor Accuracy (one-sigma)

Along-heading	0.25% + 0.1 knot
Cross-heading	0.25% + 0.1 knot
Vertical velocity	0.1% + 0.05 knot

Position accuracy is 1.3% (CEP) of distance traveled when heading reference accuracy is one degree (one sigma).

Physical characteristics of the AN/ASN-128 and SHIU are listed below.

AN/ASN-128	WEIGHT		VOLUME		POWER	MTBF
	kg	lb	cm ³	in ³	(W)	(h)
DRVS: RTA	4.76	10.5	7,752	473	12.0	12,237
SDC	5.67	12.5	9,296	564	56.6	7,744
DRVS Total:	10.43	23.0	17,048	1,037	68.6	4,742
CDU	3.18	7.0	3,676	224	30.0	3,838
Total:	13.61	30.0	20,724	1,261	98.6	2,121
SHIU	0.91	2.0	1,067	65	15.0	10,000

The AN/ASN-128 provides accurate present position, steering signals, and distance and time-to-go to 10 destinations. It uses a printed grid antenna that is self-compensating for the over-water shift. Navigation data are displayed in both UTM coordinates and latitude/longitude.

Other operational advantages include:

- Single transmit-receive antenna makes maximum use of the available aperture, which minimizes random errors - important for accurate fire control.
- Back-up modes of operation should heading, or pitch and roll, or Doppler velocity become unavailable.
- No maintenance adjustments at any support level.
- No special test equipment at the flight line: BITE localizes malfunctions to the LRU and SRU levels.
- CDU provides serial digital outputs of system navigation, guidance and other data for use by external avionics.

DESCRIPTION OF THE AN/ASN-128 AND THE SHIU

The AN/ASN-128 Lightweight Doppler Navigation system consists of three boxes that weigh only 30 pounds in total and have a predicted MTBF of over 2100 hours. The auxiliary display unit, the SHIU, weighs only 2.0 pounds, and has an MTBF greater than 10,000 hours.

The design of the AN/ASN-128 was centered on making maximum use of the existing heading and attitude sensors and complementing these with a DRVS, a computer, and suitable controls and displays.

A Doppler radar consists of an antenna, RF transmitting and receiving components, and electronics for processing the Doppler frequency shifts and converting these into the desired data format. A body-mounted antenna was selected to avoid gimbals that would otherwise add considerable size, weight and complexity. Four beams of RF energy directed downward symmetrically around the nadir are sequentially transmitted by a single antenna toward the ground, and the backscattered energy is then sequentially received and processed. Although three beams are sufficient to determine the three orthogonal components of helicopter velocity, the AN/ASN-128 uses a fourth, redundant, beam to enable self-checking of the overall reliability of the velocity measurement.

The antenna must be mounted on the helicopter's underside to enable its beam to illuminate the ground, and thus, installation and removal is generally more difficult than that of an electronics box mounted in an avionics bay. The antenna unit was therefore designed to contain the minimum amount of electronics for maximum reliability, (12,000 h) but it does include all transmit and receive RF components so that waveguide runs within the fuselage are avoided. A Singer-developed, printed-grid antenna was used to eliminate the weight and cost of waveguide arrays. Installation was further simplified by integrating the radome with the antenna. All electronics required for velocity processing are contained in a separate electronics box, the SDC, located in the avionics bay. All regulated voltages required by the RTA, SDC, CDU and SHIU are provided by a single power supply in the SDC. A single card in the SDC sequentially converts the synchro outputs of the existing heading, attitude and true airspeed sensors into digital form for use by the computer in the CDU.

All computations for the AN/ASN-128, including velocity coordinate transformations, navigation in both UTM and latitude and longitude coordinates, steering signals to 10 destinations, and BITE functions, are performed in an LSI digital computer located in the CDU. This approach minimizes the cost of the RTA and SDC when only these two boxes are used as a DRVS.

Two types of BITE test are provided: continuous and manually initiated. The CDU and most portions of the SDC and RTA are continuously tested, and the overall SDC and RTA are tested from end-to-end during the manually initiated test. All test results are displayed on the CDU; system status, the failed box, and the failed module within that box. "GO" is displayed when no failure has occurred. End-to-end test takes only 18 seconds and does not interrupt navigation and guidance computations.

The system has been in production since 1978, and to date over 3,500 LDNS systems have been ordered of which over 3,200 have been delivered to the U.S. Army for installation in the AH-1S Cobra, UH-60A Black Hawk and CH-47D Chinook helicopters. The system has been selected by, and is being manufactured for several other countries, including Germany, Spain, Greece and Australia; additional orders have been received from these countries for over 600 systems.

MISSION EFFECTIVENESS

The above section describes some of the trade-offs leading to the system architecture, performance and physical characteristics of the navigation system. The system designer, together with the operational personnel, must also examine the effectiveness of the system in performing the planned mission. To accomplish this, a multisegment mission is examined in the following section to identify the various requirements of the navigation system. Figure 4 shows the selected hypothetical, yet typical, mission scenario for a utility helicopter. This is followed by a discussion of the capability of the AN/ASN-128 to satisfy these requirements.

The postulated mission commences at the Forward Area Rearming and Refueling Point (FARRP). Coordinates in either UTM or Lat/Long format of the planned destinations or other preselected stored data should be rapidly insertable into the navigation system. A dual coordinate capability is desirable, should target handoff be required. Rapid deployment unencumbered by requirements for lengthy alignment times is essential. BITE check should be rapid yet provide high confidence. The helicopter will pass the Forward Edge of the Battle Area (FEBA) enroute to Lz Tiger for a supplies drop. Once past the FEBA, minimum radiated emission is desirable to minimize the probability of intercept. Enroute, the flight crew must be provided with present position, off-track error, distance-to-go, true ground speed and track angle. On landing at Lz Tiger, it is assumed that power would be turned off requiring that selected destination stored data not be destroyed. Should the battle environment not allow a touchdown landing, the system must display hover guidance data. Enroute to Lz Argo, enemy action may cause the helicopter to depart from its planned course. Instantaneous call-up of present position to allow reporting of the encounter, and updated guidance to Lz Argo must be provided. Having arrived over Lz Argo, which is a point of known position, a means to quickly update for any accumulated navigation error is desirable. The navigation system must be configured to continue navigation throughout the update to avoid error buildups should higher priority tasks delay completion of the update.

After updating at Lz Argo and while enroute to the next planned destination, Lz Judy, the aircraft is rerouted to an unplanned landing zone - Lz OX. The system must be capable of in-flight alternate destination entry and selection.

Having completed the mission function at Lz OX, the system should provide the ability simply to call up the prestored destination Lz Judy and provide all guidance and steering commands required to reach that Lz.

Enroute to Lz Judy a target of opportunity is observed, namely, a pontoon bridge over a stream. The ability to over-fly this target and automatically store and display its position is required. Should operational conditions prevent over-flight, it is desirable to have an offset target of opportunity storage capability.

The revised flight path to Lz Judy requires flying over a large body of water with an undulating shoreline. A navigation system that automatically provides over-water compensation eliminates the need for the flight crew to manually activate a land-sea switch. This feature is operationally desirable and, in fact, mandatory for nighttime and all-weather flying when passage over water is not detectable.

For night flying it is essential that the navigation system controls and displays be compatible with night vision goggles.

Although the AN/ASN-128 entered production as recently as 1978, the pace of an ever-increasing enemy threat both in lethality and in quantity of air forces and armor has significantly enlarged the mission requirements for future helicopters. The requirement to routinely conduct operations at Nap Of the Earth (NOE) altitudes in marginal weather under reduced visibility has placed ever-increasing demands upon the onboard avionics.

More accurate and sophisticated fire control systems coupled to LASER designators, FLIR and highly accurate attitude and velocity reference systems are required. To counter the ECM threat, improved and multifrequency communication systems have been developed and are being deployed. These demands have significantly increased the number of electronic systems on board helicopters, resulting in the problems of increased crew workload, cockpit space requirements, system cost and weight.

USE OF DIGITAL DATA BUS TECHNOLOGY

The increasing amount of electronics systems requires the transfer of large amounts of digital data at very high rates. To meet this requirement, digital signal processing techniques and architectures are being applied to interconnect sensors, displays and controls and multiple avionics microprocessors at the subsystem level.

The standard specification developed by the United States military for serial data multiplex transmission is MIL-STD-1553A/B. Military aircraft of the 1980's rely heavily on multiplex data buses for information distribution in the design of their avionics architecture. In addition, MIL-STD-1760 specifies that both expendable and nonexpendable

stores be capable of interfacing with the host aircraft via a MIL-STD-1553B digital multiplex (MUX) bus. The MIL-STD-1553 digital multiplex bus provides a means of transferring data and commands from remote avionics over the same transmission line--in this case, a shielded twisted pair. The use of a digital data bus reduces weight and improves reliability because less wire and fewer connectors are required. Digital transmission techniques provide a higher data capacity, self-check on each transmission and reduced susceptibility to electromagnetic interference.

Key to the development and utilization of the digital data bus is the MIL-STD-1553 Bus Controller and Multiplex Remote Terminal (BC/RT), which provides the interface between the aircraft MIL-STD-1553 MUX system and the onboard avionics unit in which it is contained. Singer has developed a multiplex terminal unit of universal applicability for use in avionic LRU's. The BC/RT module described in the following section utilizes high-density microcircuits and a flexible two-part architecture that is virtually independent of host-peculiar requirements, and is therefore usable in a wide range of applications with little or no modification. It is designed to be located in the host LRU since the entire module is packaged on a single card and requires less than 10 watts.

The BC/RT can operate as a controller or as a remote terminal. The mode of operation is selectable by a Master/Slave logic signal or via the MIL-STD-1553 mode command for dynamic bus control transfer.

When designated a remote terminal, the BC/RT is responsive to all MIL-STD-1553A and B command-response requirements. When designated as a bus controller, the BC/RT initiates and supervises all data exchanges over the dual redundant MIL-STD-1553 serial data bus. Data storage and retrieval at the host parallel data bus is via direct memory access.

Features include:

- a. Packaged on a single card module
- b. Performs as a Bus Controller and/or Remote Terminal capable of executing MIL-STD-1553B Mode and Illegal Command Word Processing
- c. Contains separable Word Processing and Message Processing/Microcontroller sections
- d. Word Processing section:
 - Transparent interface for MIL-STD-1553A or B compatibility.
 - Interfaces with dual redundant MUX buses. Performs all fast response front-end channel functions, including word validation, and command and data word detection and decoding.
 - Provides all necessary parallel data with the sense and control signals required by the Message Processing/Microcontroller section.
- e. Message Processing Microcontroller section:
 - Transparent architecture for interfacing with various hosts, including those with microprocessors
 - Programmable
 - 16-bit bidirectional data bus
 - 16-bit address bus
 - DMA capability to interface with host microprocessor
 - Provides memory protection and ability for indirect addressing of host main memory.

Figure 5 is a block diagram showing the two part architecture of the BC/RT and identifying the large number of functions which are packaged on the single card in less than 50 square inches by the optimum utilization of LSI technology.

Communication with the host is via a 16-bit bidirectional data bus. A 16-bit address bus and direct memory access circuits perform the functions of storage and retrieval of data to-and-from host memory. Other significant circuit elements are a 16-bit terminal bus for fast response data manipulation and transfer, and the MIL-STD-1553 Status Word and Last Command Word registers. In addition, an on-board scratchpad register and incrementor are provided.

Some examples of the application of digital data bus technology to advanced avionics are the U.S. Army's AN/ASQ-166 Integrated Avionics Control System (IACS) and the AN/ASN-137 Multiplexed LDNS.

The U.S. Army Avionics Research and Development Activity (AVRADA) developed the AN/ASQ-166 IACS which is characterized by:

- 1) Use of digital data bus to interconnect sensors, controls and displays

- 2) Integration of programmable controls and displays, and
- 3) Widespread application of microprocessors at the subsystem level.

The AN/ASQ-166 IACS provides the pilot with a means of controlling and displaying communications, navigation such as the AN/ASN-137, and Identification (CNI) equipment with a single integrated control-indicator. All CNI equipment are remotely located and their information is transferred by a dual-redundant MIL-STD-1553 data bus.

The basic IACS consists of five units - two control-indicators, one status indicator and two interface units. Figure 6 shows a block diagram of a typical IACS installation. The controlled avionics, shown at the right of Figure 6, are remotely located in the avionics bay of the aircraft. The present IACS system can accommodate up to ten radios and associated communications, security equipment and the AN/ASN-137 Doppler Navigation Set.

The Primary Control-Indicator integrates the control and display functions of all ten CNI and associated security equipment and the Doppler Navigation Radar. This integration is accomplished by time-sharing both controls and displays of a large number of previously unrelated systems. Time-sharing of controls and displays conserves panel space by using fewer controls and displays to accomplish the necessary avionics control functions.

The Secondary Control-Indicator provides a minimum capability for emergency situations.

The U.S. Army has embarked on a development program to extend the IACS integration concept from the CNI functions to the remaining cockpit functions. This program has been designated the Army Digital Avionic System (ADAS). IACS is a subset of the ADAS system. Elements of ADAS have been installed in a UH-60A Helicopter Systems Testbed for Avionics Research (STAR) and are undergoing flight test. The AN/ASN-137 is one of the avionic systems installed in the ADAS installation in STAR along with the IACS.

DESCRIPTION OF THE AN/ASN-137

Another example of the use of advanced digital data bus technology that has been developed by the U.S. Army is the Singer AN/ASN-137 Doppler Navigation System.

Overall design of the AN/ASN-137 was based upon two requirements:

- Provide a system that has MIL-STD-1553 interface, drives a Projected Map Display System (PMDS) and can operate with an integrated display such as IACS,
- Retain as much of the AN/ASN-128 hardware design as possible to minimize development costs and to have maximum commonality with the AN/ASN-128 LRU's, SRU's, and logistics support equipment.

The changes required for the AN/ASN-137 to meet the above requirements do not affect the Doppler radar portion of the AN/ASN-128. Therefore the Receiver-Transmitter-Antenna (RTA) and a major portion of the Signal Data Converter (SDC) need not be changed. Avoiding RTA changes simplifies retrofit of the AN/ASN-137 into helicopters already containing the AN/ASN-128.

The need to provide navigation, guidance, and velocity data when operating with an integrated display such as IACS, instead of the AN/ASN-128 CDU, required the computer and memory in that CDU to be transferred to the SDC.

A simplified CDU was developed and is available for those applications of the AN/ASN-137 where IACS is not provided. A major design feature of the simplified CDU is that its operation is the same (except for PMDS related operations) as that of the AN/ASN-128 CDU to minimize flight crew retraining and chance of confusion when flight crews operate different helicopters.

The MIL-STD-1553 A and B interface was added to the expanded SDC to provide a digital communication capability between the AN/ASN-137 and other avionics in the helicopter. Locating the MIL-STD-1553 interface in the expanded SDC was also desirable since its design would be simplified by having direct access to the computer bus. To handle the increased computational load of the MIL-STD-1553 interface and other avionics (i.e., projected map drive equations, offset targeting, etc.), the AN/ASN-128 program memory was increased from 8,000 to 16,000 16-bit words while still providing about 3,000 words of spare memory. The random access memory was increased from 500 to 1,000 16-bit words providing about 200 spare words.

To provide for a variety of board sensors and displays, a "Program Plug" has been added to the outside of the SDC. Each helicopter designed to accept the AN/ASN-137 will have a captive cable that connects to the Program Plug, and reprograms the interface within the AN/ASN-137 to be compatible with the configuration in that particular helicopter.

A block diagram of the overall AN/ASN-137 system is shown in Figure 7. Some of the key system features are listed below:

- Provides MIL-STD-1553 two-way multiplexed bus interface
- Can be controlled either by an integrated control/display system such as IACS, or a simplified version of the AN/ASN-128 CDU
- RTA unchanged for ease of retrofit
- Program plug enables operation in different helicopter avionics configuration without changes
- Drives AN/ASN-99 Projected Map Display System (PMDS).

AN/ASN-137	Weight		Volume		Power (W)	MTBF (h)
	kg	lb	cm ³	in ³		
DRVS: RTA	4.76	10.5	7,752	473	12.0	12,237
SDC	7.04	15.5	9,896	603	75.0	3,723
DRVS Total:	11.80	26.0	17,648	1,076	87.0	2,855
Simplified	2.82	6.2	3,676	224	30.0	4,584
CDU						
Total:	14.62	32.2	21,324	1,300	117.0	1,760
SHIU	0.91	2.0	1,067	65	15.0	10,000

The AN/ASN-137 completed its engineering development phase in late 1980, and was integrated by Singer with the AN/ASQ-166 IACS, and the AN/ASN-99 Projected Map Display in early 1981. The system started U.S. Army flight test in mid-1981, and has been in production since 1982.

SYSTEM CONSIDERATIONS FOR HIGH ACCURACY APPLICATIONS

IMPROVED NAVIGATION ACCURACY

Figure 9 shows the position error of the AN/ASN-128 or AN/ASN-137 Doppler Navigation Systems when a typical helicopter magnetic compass is used to provide heading. An error of one degree, (one-sigma), is characteristic of such compasses, resulting in a CEP of approximately 1.3% of distance traveled. For this system configuration, the position error after 50 kilometers of flight is 650 meters, as shown by the solid line in the figure. The dotted line shows the error build-up when an improved heading reference, with a heading error of 0.5° (one-sigma) is used. This accuracy can be achieved by careful compensation of magnetic compass deviation error and by data stabilization of helicopter pitch and roll effects on heading.

The increasing performance requirements of advanced attack and scout type helicopters has led to the need for a heading reference with 0.25° (one-sigma) accuracy and thus requires the use of a new system approach.

THE SKH-3700 ATTITUDE HEADING AND NAVIGATION SYSTEM (AHANS)

In response to the need for a more accurate heading reference, Singer has developed the SKH-3700 AHANS. The AHANS is a single-box system that utilizes strapdown inertial sensor technology to provide very accurate heading, attitude, acceleration and angular rate data via a MIL-STD-1553 MUX I/O. The AHANS combines velocity data from the AN/ASN-128 or AN/ASN-137 with its own inertial sensor data in an optimum Kalman filter mechanization. The combined system has a heading accuracy of better than 0.25° (one-sigma).

The AHANS consists of a strapdown inertial sensor block, a powerful computer (SKC-3121) programmed in a high order language, I/O, and power supply. Two, two-degree-of-freedom, "dry" CONEX gyros, and three, "dry", pendulous accelerometers make up the sensor block. Figure 10 shows a block diagram of the AN/ASN-128 operation with the AHANS. Figure 11 shows the AHANS LRU, typical electronics cards, the sensor block, the gyro and accelerometer.

The Kalman filter in the computer software is a 24-state mechanization. It models both system navigation parameters and combined Doppler/AHANS component error sources. Through the optimum combination of inertial sensor data, Doppler velocity data, and external position observations (when available), the AHANS refines estimates of vehicle position, velocity, and attitude during flight and continually calibrates Doppler/inertial component error sources.

This mechanization provides a rapid reaction "scramble" or in-flight startup capability so important in military operations, without impacting navigation accuracy later in the flight. In addition, improved navigation is available during periods of back-up mode operation, such as Doppler in memory.

A prototype version of AHANS was flight tested by the U.S. Army, in operation with an AN/ASN-128. Navigation results were excellent, with position errors less than 400 meters after 2 hours of flight. Heading errors averaged less than 0.25° (one-sigma) during this time.

The AHANS with its powerful digital computational capability, when interfaced via a MIL-STD-1553 MUX bus with the Doppler, and controlled by IACS, will provide not only navigation, but also velocity, heading, attitude, acceleration and angular rates with the accuracies needed for future requirements.

EXTERNALLY REFERENCED SYSTEMS

Another level of system performance improvement and operational capability can be obtained by combining the AN/ASN-128 or the AN/ASN-137 with externally-referenced systems, such as the Global Positioning System (GPS), Position Location and Reporting System (PLRS) or the Joint Tactical Information Distribution System (JTIDS). As was shown in Figure 9, the position error of a Doppler radar navigation system increases with distance traveled. The position error of an externally-referenced system is essentially bounded and predictable. A combined Doppler/position sensing system has several important advantages. The position error is bounded to that of the position sensing system; errors in the Doppler navigation data (velocity and heading) are calibrated by the position sensing system, and finally, accurate navigation and guidance data continue to be available to the flight crew even when operation of the position sensing system is interrupted for any reason, including intentional interdiction by the enemy. As future position sensing systems become mature, one can expect increased interest in combining these systems to obtain more optimum navigational accuracy and reliability.

TECHNOLOGY ADVANCES TO MEET NEW REQUIREMENTS

As electronic components decrease in size, weight and power requirements, the demands for increased capability and functions tend to increase to fill the available space created. For example, the Doppler radar echo contains altitude (distance above the ground) information in the phase of its modulation. As microprocessor and VLSI technology develops, it will be possible and desirable to add an altimeter function to the Doppler navigation radar, basically without penalty.

CRT displays, when used in combination with microprocessors, are providing new versatility in the amount, format and quality of the information which can be provided to the pilot. The use of multi-color CRT's adds the dimension of color to the data, allowing highlighting of critical data on the CRT face. The advent of flat panel displays will save volume, weight and power over current CRT's.

The increase in memory capacity, which Magnetic Bubble Memories and VLSI offer, will allow digital storage of maps, which can be displayed on a CRT or flat panel together with position, track-made-good, and track-to-go superimposed from the onboard navigation system.

As mentioned above, the progress in circuit miniaturization will produce major changes in the size, weight, power consumption and capabilities of avionic hardware. Microprocessors are being used in almost all signal processing applications, replacing bulky analog devices such as filters and inductors. Discrete microwave circuits using waveguide interconnections are being replaced by printed circuit equivalents (called "Stripline" or "Microstrip"), with integrated active devices embedded in the printed circuitry.

Singer has been in the forefront in applying these technologies to its product lines, especially on printed microstrip antennas. A printed microstrip Doppler radar antenna has been developed to replace the AN/ASN-128 printed grid antenna. It is only 0.030 inches (0.8 mm) thick and can be applied to conform to an aircraft surface. This new version weighs 2 lb less and requires only 40% of the present volume.

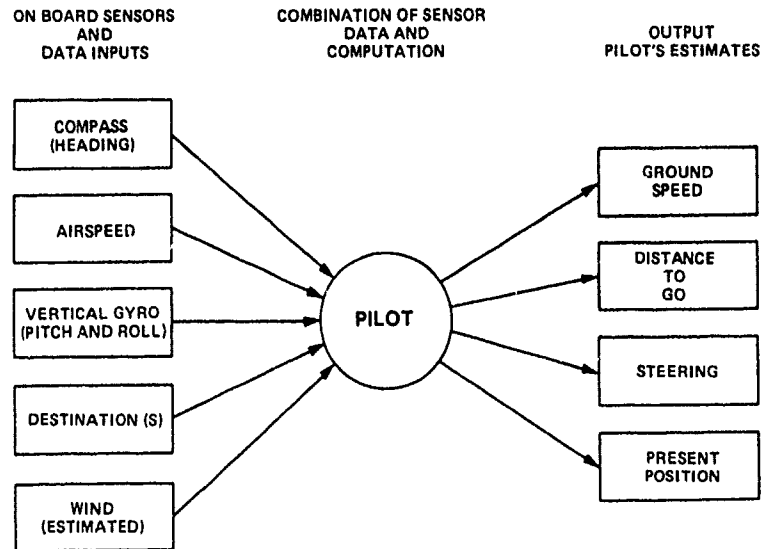
Figure 12 graphically illustrates the benefits of technological breakthroughs to avionic systems. The criterion used here was the weight of Doppler navigation radar systems produced since 1950. The dramatic decrease in weight between 1950 and 1960 is primarily due to replacing vacuum tubes with transistors. Between 1960 and 1970 the weight decreases were accomplished by improvements in antenna technology, resulting in smaller antennas and much simpler antenna stabilizing gimbals. In addition, improvements in RF hardware permitted the replacement of pulsed high power transmitters by more efficient solid state sources using FM-CW modulation techniques. Since 1970, further weight reductions were achieved by replacing mechanical antenna stabilization with digital data stabilization, replacing slotted waveguide antennas with printed antennas, and by the extensive use of digital logic and MSI/LSI components. Each reduction in weight was accompanied by a reduction in cost and size.

The current rapid pace of developments in VLSI, microprocessors, microwave integrated circuits and advanced digital signal processing techniques will afford opportunities for even greater degrees of avionic sensor integration.

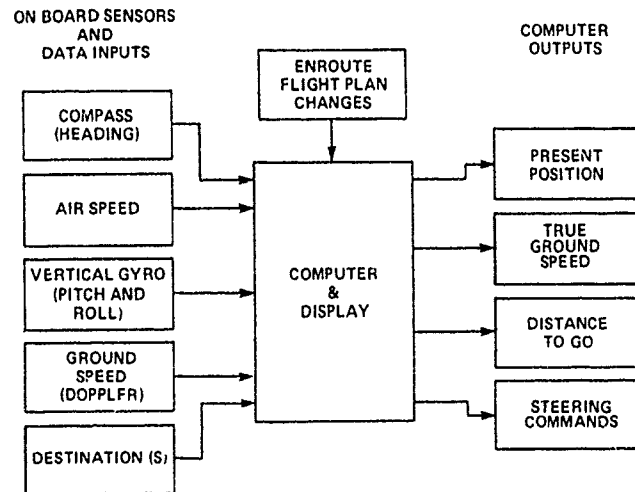
By virtue of these costs and size reductions, we will see operational capabilities currently supplied in sophisticated fixed wing aircraft extended to future helicopters and drones.

REFERENCES

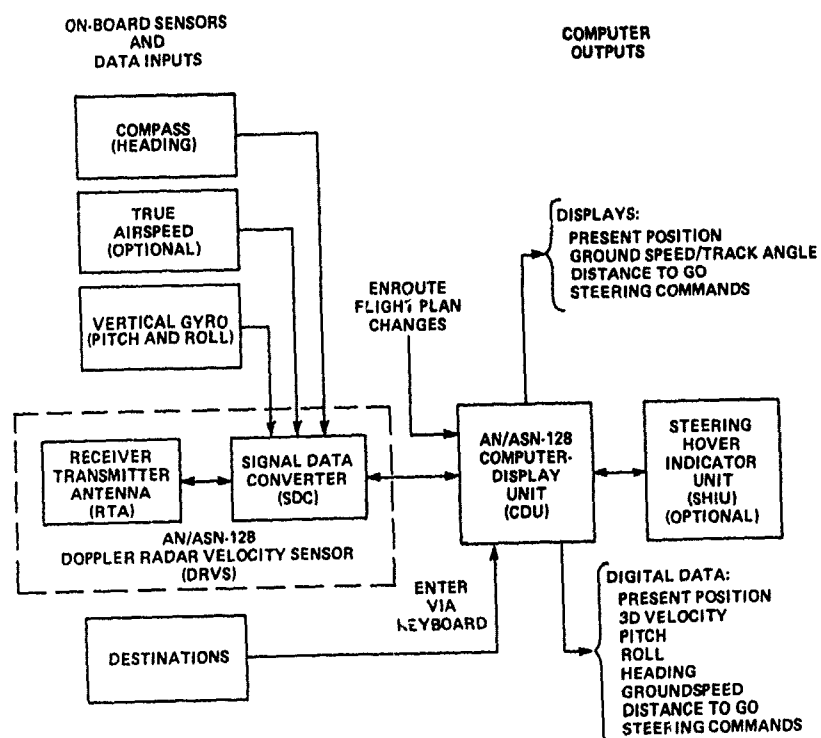
1. Galanti, C.J. and Santanelli, A. S., "AN/ASQ-166: An Approach to Integrated Avionics Systems," presented at the 3rd Digital Avionics Systems Conference, Fort Worth, Texas, November 1979.
2. Dasaro, Dr. J.A. and Elliott, C.T., "Integration of Controls and Displays in U.S. Army Helicopter Cockpits," at the 32nd Symposium of the Guidance and Control Panel, Stuttgart, Germany, May 1981.
3. Sanducci, A.F., "MIL-STD-1553 Bus Controller and Multiplex Remote Terminal" presented at IEEE 1981 National Aerospace and Electronics Conference, Dayton, Ohio, May 1981.
4. Buell, H., "Doppler Radar Systems for Helicopters," published in the Journal of The Institute of Navigation, Vol. 27, No. 2, Summer 1980.



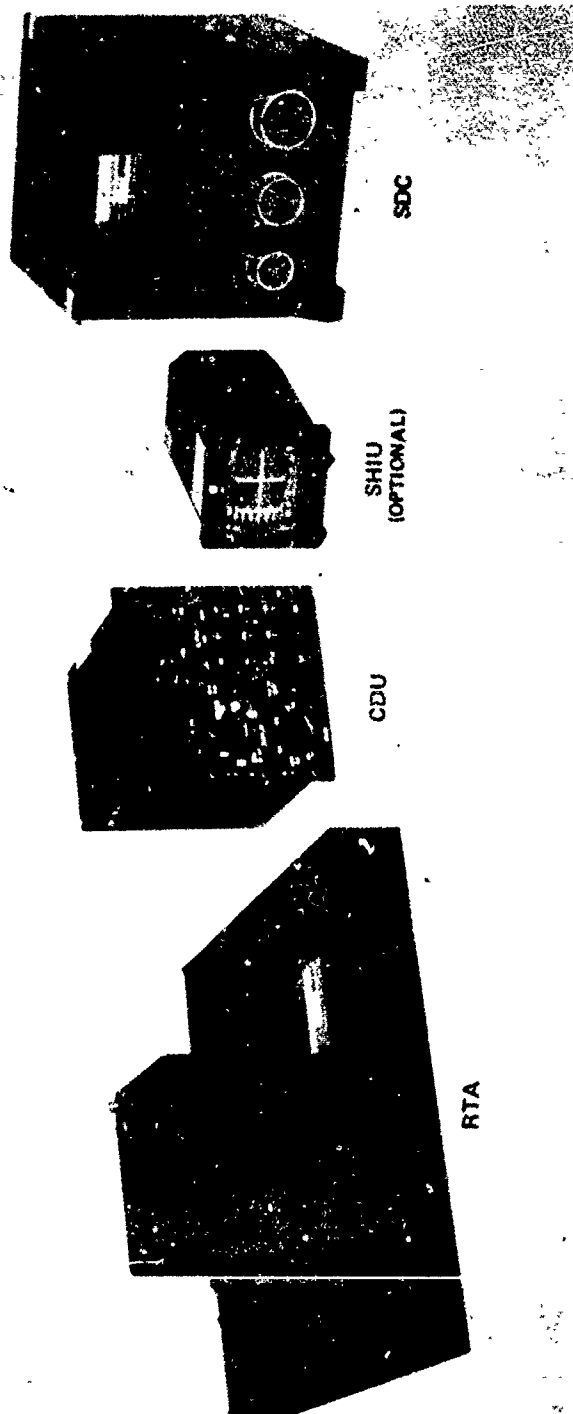
NAVIGATION USING MANUAL INTEGRATION OF INPUTS
FIGURE 1A



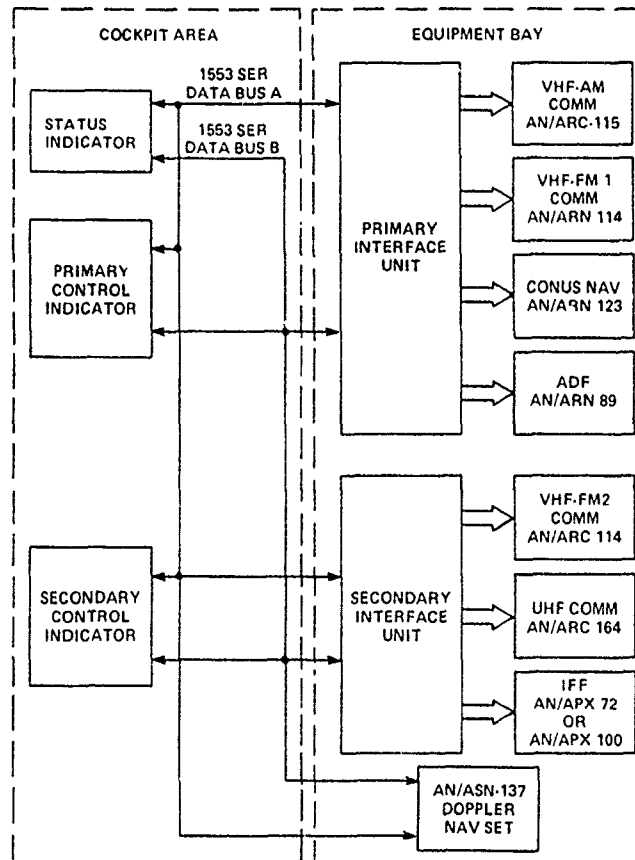
NAVIGATION USING COMPUTER INTEGRATION OF INPUTS
FIGURE 1B



AN/ASN-128 BLOCK DIAGRAM
FIGURE 2

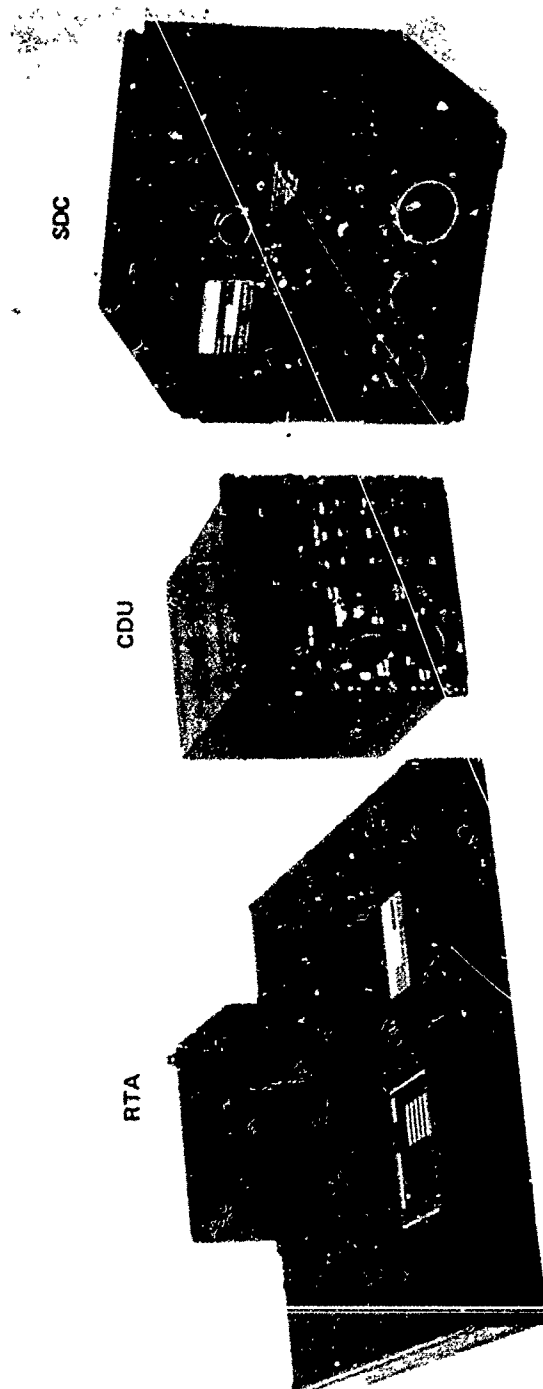


AN/ASN-128 LIGHTWEIGHT DOPPLER NAVIGATION SYSTEM
FIGURE 3

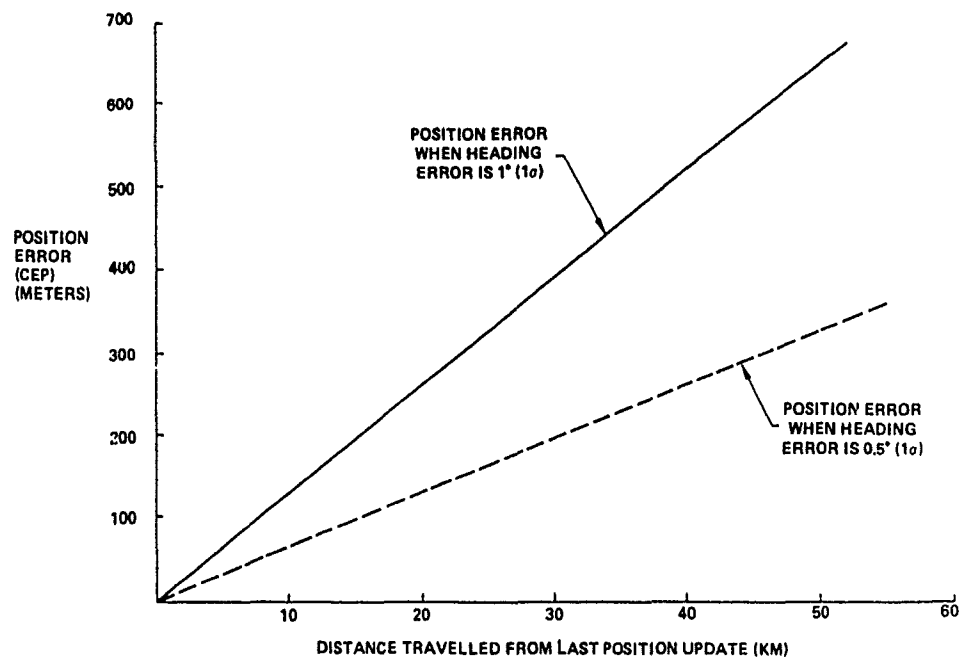


INTEGRATED AVIONICS CONTROL SYSTEM, AN/ASQ-166
FIGURE 6

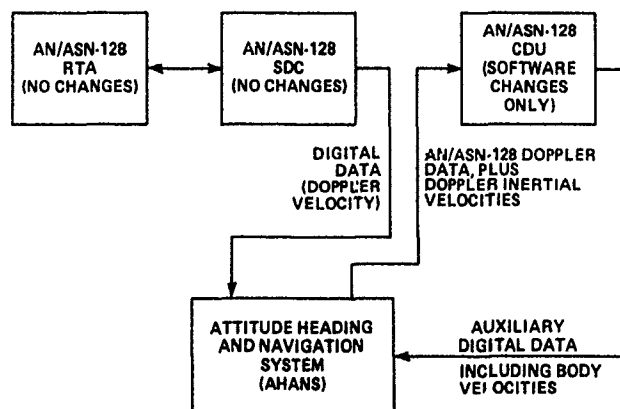
AN/ASN-137 DOPPLER NAVIGATION SYSTEM BLOCK DIAGRAM
FIGURE 7



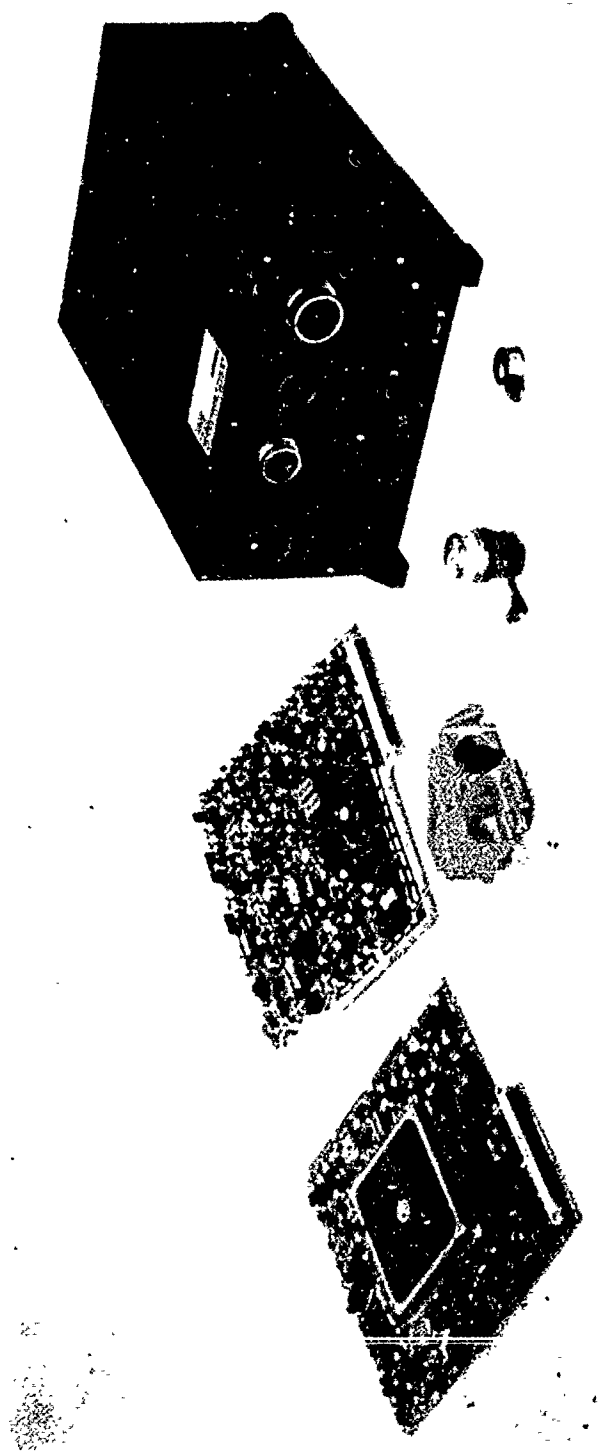
AN/ASN-137 DOPPLER NAVIGATION SYSTEM
FIGURE 8



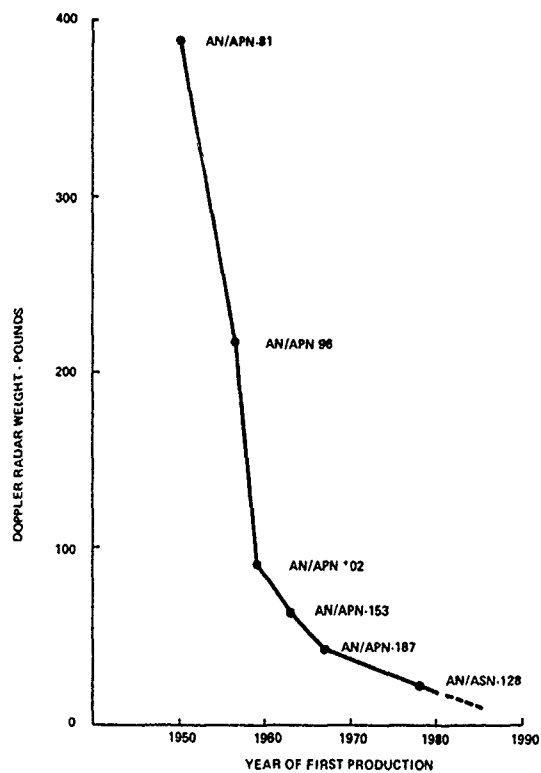
POSITION ACCURACY OF AN/ASN-128 OR AN/ASN-137
FIGURE 9



AN/ASN-128 AND AHANS BLOCK DIAGRAM
FIGURE 10



ATTITUDE HEADING AND NAVIGATION SYSTEM (AHANS)
FIGURE 11



IMPACT OF TECHNOLOGY ADVANCES ON DOPPLER RADAR WEIGHT
FIGURE 12

PART II

**NAVSTAR/GPS
(Global Positioning Satellite) Systems**

TECHNIQUES FOR AUTONOMOUS GPS INTEGRITY MONITORING

by

Professor Bradford W. Parkinson and Penina Axelrad
Stanford University
W.W. Hansen Labs
GP-B, Via Palou
Stanford, CA 94305
United States

SUMMARY

The use of GPS for navigation critical applications such as aircraft non-precision approach or harbor and river crossings requires the navigation data to be both extremely accurate and extremely reliable. This paper outlines various approaches to GPS integrity checking, and describes a method for user autonomous satellite failure detection and isolation (D/I). The test statistic for the D/I algorithm is the range residual parameter based on six or more satellites in view. The nominal pseudo range measurement errors are modeled as normally distributed with mean in the range of -5m to +5m and standard deviation of 0.4m based on experiments conducted at Stanford. The theoretical statistical distribution of the range residual is given. Monte Carlo simulations present results of applying the algorithm to measurement sets containing a biased measurement. With a 100m biased measurement present successful detection is achieved 99.9% of the time, and successful detection and isolation is achieved 72.2% of the time. The user is always aware when isolation is not possible. User positioning errors resulting from application of the algorithm are always the same or better than the all in view solution.

1.0 INTRODUCTION

As the Global Positioning System (GPS) nears its operational capability, the number and diversity of the potential applications being suggested is increasing at an accelerating pace. Already GPS has become a standard tool for surveying. Experimental terrestrial, marine, and airborne navigation is also being done during the limited periods of adequate satellite coverage. There is also considerable interest within the potential user community in applying GPS to a variety of high precision navigation problems such as river and harbor navigation, aircraft non-precision approach, real time or kinematic surveying, and spacecraft orbit determination. These applications not only require highly precise navigation data, but they need it all the time. Thus arises the growing interest in GPS integrity monitoring. Significant research efforts addressing integrity issues have been going on since about 1980 [1].

The objective of all integrity verification schemes is to ensure the accuracy of a user's navigation solution in the presence of degraded measurements. This function is extremely important during critical mission phases such as an aircraft in final approach, a ship traversing a narrow passageway, or a low earth orbit satellite during orbit injection or final orbit trim.

A user's navigation solution may degrade because of a GPS satellite failure, poor geometry, selective availability (SA), local conditions, or a user receiver failure. Most integrity monitoring schemes directly address the problem of a GPS satellite clock failure or degradation of the broadcast ephemeris. In general these techniques are also useful for flagging a particularly bad combination of SA errors and user/satellite geometry. It is more difficult to protect against the influences of local multipath reflections and tropospheric delay variations because they vary substantially in magnitude and character. Research work is being conducted at Stanford and elsewhere to uncover the structure of errors caused by multipath interference. Receiver failures, such as uncalibrated interchannel biases, can generally be identified within the receiver by periodically tracking the same satellite on more than one channel.

The next section introduces the various approaches to integrity verification which have been suggested. Following this receiver autonomous monitoring is expanded further in Section 3. Section 4 presents results from experiments conducted to establish the true structure of GPS pseudo range measurements. Section 5 outlines the least squares technique and describes the characteristics of the key test statistic which is the basis for the Stanford approach to failure detection and isolation. Section 6 presents the straightforward monitoring procedure. Section 7 shows results from extensive Monte Carlo simulations, and discusses the results and their implications. Section 8 provides a summary of what has been covered, and suggests future

enhancements and applications.

2.0 APPROACHES TO INTEGRITY MONITORING

There are three basic ways for a user to determine if a failure has occurred: 1) wait for official notification via the satellite navigation message; 2) wait for notification from an integrity monitoring station; 3) monitor the navigation solution integrity autonomously.

2.1 Official Notification

Official notification of a degraded satellite measurement is provided by the SV health data in the GPS telemetry word (TLM). Two types of health data are contained in the fourth and fifth subframes of the TLM broadcast by each satellite [2]. A three-bit navigation data health indicator describes the status, and in case of error, the type of problem for the transmitting satellite. Each satellite TLM also contains a one-bit status code for all satellites in the constellation. Health status updates are routinely performed along with satellite ephemeris uploads or when a change of status is necessary [2].

The main problem with using the health status indicators for integrity monitoring is that the response time to a satellite failure may range from 20 minutes to several hours. The actual delay depends on the location of the satellite with respect to one of the four GPS monitor stations, and the task loading of the GPS Master Control Station at the time of failure [3]. Official notification only provides satellite health status and thus is also not helpful in the case of poor geometry, severe SA, or localized sources of error.

2.2 Monitor Station

The Federal Aviation Administration (FAA) has been the primary group interested in the development of independent integrity monitor stations. Since 1980, a series of studies by the MITRE Corporation has been sponsored by the FAA to investigate and define a system of ground monitor stations. These stations located at surveyed positions would monitor satellite signal integrity and notify users within a small area of satellite failures within 15 sec [4,5]. Differential GPS stations providing range error information to nearby users could also serve as effective monitor stations. A joint study by Ford Aerospace, STI, and the Aerospace Corporation investigated an alternative approach in which integrity data is uploaded to a geostationary satellite and broadcast to all users on an integrity channel [6].

The ground monitor stations offer a distinct advantage for users in the vicinity of airports where continuous integrity monitoring is crucial. The disadvantage is that the cost of a ground station will probably prohibit its use except in high density areas.

2.3 Autonomous Monitoring

In many instances it may be in a user's best interest to provide self protection against GPS navigation system degradations. Autonomous monitoring techniques can be divided into two categories. The first uses a complementary navigation sensor to verify GPS measurement or navigation solution integrity. Integrated INS/GPS, AHRS/GPS, and Radar/GPS are a few examples of such systems [7]. Clearly they can be set up to protect against both single satellite failures and systematic errors in the GPS system or the user receiver. The added cost of the secondary navigation sensor varies tremendously as does the reliability of the navigation data. Few systems are sufficiently accurate to be used as an effective check on GPS.

The second methodology for autonomous integrity checking uses GPS system redundancy to monitor measurements or position solutions. The recent announcement by the GPS Joint Program Office that it supports the return to a 24 satellite baseline constellations is quite a boost to proponents of monitoring schemes based on satellite redundancy [8]. With the full 24 satellite constellation most terrestrial users will have a minimum of 6 satellites in view. Since only 4 satellites are required to solve for a user's position and clock bias, under certain conditions the availability of the additional measurements can be used for a consistent or integrity check. Over the past few years, a variety of such schemes have been proposed.

This is the main topic of this paper and will be developed in detail in the following sections.

3.0 RECEIVER AUTONOMOUS INTEGRITY MONITORING

Receiver autonomous integrity monitoring (RAIM) is the name coined by the FAA for techniques which rely only on redundant GPS satellite measurements for failure detection [9]. The following sections elaborate on the goals, constraints, and techniques particular to this class of autonomous GPS user.

3.1 Objectives

As previously discussed, the overall goal of integrity monitoring is to ensure the accuracy and reliability of a navigation system during critical mission phases. However, more specific objectives for autonomous monitoring can be identified.

Solution monitoring is currently receiving the most attention in the GPS integrity literature. This truly is the issue of primary concern to a large class of users, for example, an aircraft about to execute a non-precision approach. Solution monitoring methods set a 2-D or 3-D "protection limit" which indicates the position offset the user can tolerate without considering a navigation system failure.[9]

Measurement monitoring is concerned with identifying an actual problem with a GPS satellite measurement. It is similar to the type of monitoring done at the ground based stations; however the autonomous algorithm does not have access to outside information.

Failure isolation is merely an extension of measurement monitoring. In this case the objective is not only to detect a satellite failure but also to identify which satellite has failed. Once this has been accomplished the user can take the appropriate action to improve his navigation information.

In addition to striving for reliable detection, any RAIM technique must also be compatible with the onboard navigation system capabilities. Thus a tradeoff between performance and computational burden is essential.

3.2 Autonomous Monitoring Techniques

Numerous algorithms for autonomous monitoring have been suggested by researchers such as Y.C. Lee of MITRE Corporation [10-12], G. Brown, P.Y.C. Hwang, and P. McBurney of Iowa State University [13-15], A. Brown of Navstar Systems Development [16], and B.W. Parkinson and P. Axelrad of Stanford University [17-19]. The majority can be described as "snapshot" approaches because they use single set of GPS measurements collected simultaneously. The researchers at Iowa State University have also investigated the use of Kalman filtering as an alternative to the snapshot solution.

The advantage of a Kalman Filtering approach is that it allows you to incorporate more information in making an integrity decision. However, care must be taken to ensure that failure model assumed in the filter does not preclude failure identification under a different failure mode. Significant changes in measurement residuals can be used to flag a potentially failed satellite. Ramp type errors in a satellite clock or ephemeris prediction are more difficult to detect because they can slowly lead the solution away from the user's true position.

The snapshot approaches have a number of common characteristics. Lee pointed out that failure hypothesis testing in either the measurement or the solution domain is mathematically equivalent [10]. The methods proposed by Lee, R.G. Brown, and A. Brown form a test statistic based on a subset of the n available satellites. In the case of Reference 14 the maximum separation distance between position solutions formed from $n-1$ measurements is the test statistic. This type of technique is only reliable when the satellite geometry is favorable for all subsets of $n-1$ satellites. If any of the $n-1$ satellite subsets has a high GDOP or HDOP value, the test statistic may indicate a significant position error even though the all in view solution is quite accurate.

The technique which we have developed at Stanford is geared toward measurement detection and isolation (D/I). It is based on a single all in view solution for detection, and subset solutions for failed satellite isolation. The test statistic, known as the residual parameter, is the normalized root sum square of the pseudo range measurement residuals.

4.0 EXPERIMENTAL DETERMINATION OF RANGING ERROR STRUCTURE

An accurate model of both nominal and degraded measurement errors is essential for the design and evaluation of any integrity verification scheme. Systematic experiments aimed at uncovering the pseudo range measurement error structure were conducted at Stanford University using a state of the art GPS surveying receiver. The following sections describe the experiments and the results obtained.

4.1 Equipment

The GPS receiver used for the ranging error tests is a Trimble Navigation 4000S Model Surveying Receiver. This 5-channel set tracks up to 5 satellites and provides the option of using doppler information in addition to the raw pseudo range measurements [20]. A Trimble micro-strip antenna was mounted atop the High Energy Physics Lab at Stanford University, at a reference point surveyed by Geophysical Survey Inc [21]. This type of antenna has been found to reduce the occurrence of multipath interference. An IBM PC

XT was used to control the receiver using Trimble data collection software, and for data storage and

analysis. Information from each of the 5 channels was stored every 15 sec.

4.2 Technique

The 4000S has the capability to report the predicted range and pseudo range measurements to each of 5 satellites in view. The pseudo range measurement is based on the signal transit time from satellite to user corrected for satellite clock errors, ionospheric and tropospheric delays, relativistic effects, satellite clock offset, group delay, etc. [20]. It may also be adjusted using integrated doppler if the doppler aiding option on the receiver is selected. This pseudo range measurement contains errors due to the satellite and user clock model inaccuracies, propagation link delay model errors, multipath interference, receiver noise and interchannel biases.

The predicted range, which is called "distance" in the Trimble reference manual, is based on the receiver's current estimate of position and the satellite ephemeris data provided in the navigation message. Normally the receiver estimates its position using the combination of 4 satellites in view with the lowest GDOP value. To investigate the independent measurement errors we suppressed this function of the receiver and provided it with the known antenna location:*

Latitude 37:25.6861 N , Longitude 122:10.4690 W , Height 6.94 m

The only significant error in the calculated distance is due to the satellite ephemeris.**

The true ranging errors are then isolated by computing the difference between the pseudo range, ρ , and the distance plus clock offset, $d + b$.

$$\rho = d_0 + b - \epsilon_c - \epsilon_p - \epsilon_m - \epsilon_r \quad (1)$$

$$d = d_0 + \epsilon_a \quad (2)$$

$$\begin{aligned} \epsilon &= d + b - \rho \\ &= \epsilon_a + \epsilon_c + \epsilon_p + \epsilon_m + \epsilon_r \end{aligned} \quad (3)$$

where

ϵ_a = ephemeris errors

ϵ_c = user clock bias estimation errors - uncorrected satellite clock errors

ϵ_p = propagation link errors, i.e. uncorrected iono and tropo delays

ϵ_m = multipath errors

ϵ_r = receiver errors

4.3 Experimental Results

Pseudo range measurements were collected on various days in August and September 1987. The results are classified into three categories,

- 1) Nominal Doppler Aided are doppler aided pseudo range measurements with no loss of signal lock,
- 2) Nominal Unaided are pseudo range measurements based on code only, satellite above 15 degrees elevation, health status good.
- 3) Degraded are all other types of measurements including aided measurements during repeated loss of signal lock, measurements made to satellites near the horizon, measurements made to satellites that have failed for any reason.

The objective of the integrity verification algorithm is to identify the presence of degraded measurements (category 3) within a set of nominal measurements.

Tables (1,2) summarize the results for category 1,2 measurements respectively, and Figures (1,2) illustrate typical pseudo range error traces. We originally anticipated that SV8 would serve as a good example of a failed satellite since it is operating on a quartz crystal clock. However, the experimentally determined ranging errors to SV8 were not substantially different from typical nominal unaided measurements.

* Antenna location based on geodetic survey accurate to at least the sub-meter level. [21]

** When the position estimation function is suppressed, the distance computed by the 4000S does not ac-

count for earth rotation. The necessary correction was applied after the data was collected.

Table 1. Nominal Aided Range Error Statistics (meters)

DATE	SV03		SV06		SV09		SV11		SV12		SV13	
	mean	σ	mean	σ	mean	σ	mean	σ	mean	σ	mean	σ
AUG 19	3.37	0.42	3.90	0.39	*	*	-4.38	0.16	-2.38	0.17	-0.52	0.15
AUG 26a	*	*	0.08	0.15	-2.79	0.38	-0.14	0.12	0.66	0.15	2.19	0.18
AUG 26b	-2.65	0.33	*	*	-1.54	0.43	*	*	2.01	0.23	2.18	0.15
SEP 1b	1.38	0.36	*	*	0.77	0.16	-3.92	0.11	-2.63	0.21	4.39	0.32
SEP 2a	*	*	0.59	0.16	0.95	0.11	-1.42	0.23	-2.71	0.16	2.59	0.13
SEP 2b	4.13	0.27	*	*	-0.26	0.40	*	*	-2.84	0.39	-1.03	0.59
SEP 9b	0.16	0.18	*	*	3.12	0.18	-3.02	0.33	-2.74	0.28	2.48	0.22
AVERAGE	1.28	0.31	1.52	0.24	0.04	0.28	-2.58	0.19	-1.52	0.23	1.75	0.25

Note: The sum of the means for each session should be approximately 0m because a bias common to all measurements is absorbed in the clock estimate.

Table 2. Nominal Unaided Range Error Statistics (meters)

DATE	SV03		SV06		SV09		SV11		SV12		SV13	
	mean	σ	mean	σ	mean	σ	mean	σ	mean	σ	mean	σ
AUG 17	2.06	2.56	-2.30	2.29	*	*	-1.08	1.84	1.77	1.69	-0.45	1.28
SEP 5a	*	*	3.18	1.93	-0.06	1.29	-2.42	1.70	-1.64	1.66	0.94	1.29
SEP 5b	-1.39	2.07	*	*	2.21	1.85	-2.10	3.24	-0.74	1.85	2.01	2.11
SEP 10a	*	*	3.29	1.96	-0.69	1.34	-1.48	1.67	-1.96	1.66	0.83	1.50
SEP 10b	1.63	1.91	*	*	1.94	1.88	-2.04	3.47	-1.23	1.83	-0.30	2.03
AVERAGE	0.77	2.18	1.39	2.06	0.85	1.59	-1.82	2.38	-0.76	1.74	0.61	1.64

Note: The sum of the means for each session should be approximately 0m because a bias common to all measurements is absorbed in the clock estimate.

These results led us to formulate the following measurement error models:

- 1) Nominal Doppler Aided- normally distributed random variable with mean ranging from -5m to + 5m and standard deviation of 0.4 m.
- 2) Nominal Unaided- normally distributed random variable with mean ranging from -5m to + 5m and standard deviation of 4.0 m.
- 3) Degraded- normally distributed with σ of 0.4 m for aided and 4.0 m for unaided, and a mean greater than 5 m.

The objective of the integrity verification algorithm then, is to reliably detect measurements for which the bias is greater than the nominal mean measurement error.

5.0 BACKGROUND THEORY

This section provides background on the least squares estimation process necessary for GPS integrity checking. This brief explanation may be a review for some, but certain topics which are critical to the development of our algorithm, such as the idempotent properties of the projection matrix, are likely to be unfamiliar to some readers. This review will also serve to ensure that readers already familiar with the concepts understand the notation and terminology used throughout this article.

Following this is a description of the distribution of the normalized error sum of squares which forms the basis for the range residual parameter used as a test statistic in the D/I algorithm.

5.1 Least Squares Estimation in GPS

The pseudo range measurement equation is given by,

$$\rho_i = D_i - [e_i^T \ 1] x - \varepsilon_i \quad (4)$$

where

ρ_i - pseudo range measurement to SV_i ;

D_i - projection of the vector from the earth center to SV_i onto the line of sight from the user to SV_i ;

e_i - unit vector along line of sight from user to SV_i ;

x - (4×1) matrix consisting of the vector from the earth center to the user and the user clock bias ;

ε_i - normally distributed measurement error $\sim N(\mu_i, \sigma_i)$;

The pseudo range measurements to the n satellites in view are combined in the following matrix equation,

$$y \equiv (D - \rho) = Gx + \varepsilon \quad (5)$$

where

$$D = \begin{bmatrix} D_1 \\ \vdots \\ D_n \end{bmatrix} \quad \rho = \begin{bmatrix} \rho_1 \\ \vdots \\ \rho_n \end{bmatrix} \quad G = \begin{bmatrix} e_1^T & 1 \\ \vdots & \vdots \\ e_n^T & 1 \end{bmatrix} \quad \varepsilon = \begin{bmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

The least squares estimate of x , denoted by \hat{x} , is then given by,

$$\hat{x} = (G^T G)^{-1} G^T y \quad (6)$$

Based on \hat{x} an estimate of y can be formed;

$$\hat{y} = G\hat{x} = Py \quad (7)$$

where

$$P = G(G^T G)^{-1} G^T$$

P is commonly referred to as the "projection matrix" or the "hat matrix" [22,23].

The matrix of range residual errors, $\hat{\varepsilon}$, is the difference between the predicted and measured ranges. This ends up being equivalent to the difference between y and \hat{y} .

$$\hat{\varepsilon} = \begin{matrix} (D - G\hat{x}) \\ \text{\textit{predicted}} \end{matrix} - \begin{matrix} \rho \\ \text{\textit{measured}} \end{matrix} \quad (8)$$

$$\begin{aligned} \hat{\varepsilon} &= D - \rho - G\hat{x} \\ &= y - \hat{y} \\ &= (I - P)y = (I - P)\varepsilon \end{aligned} \quad (9)$$

The matrices P , and $(I - P)$, are $(n \times n)$ idempotent matrices with

$$\text{trace}(P) = \text{rank}(P) = \text{rank}(G) = 4$$

$$\text{trace}(I - P) = n - 4 \quad [24].$$

An orthogonal transformation, K can be found which diagonalizes $(I - P)$ to an $n \times n$ matrix with $n - 4$ diagonal elements equal to "1" and 4 diagonal elements equal to "0". The sum of the squares of the range residual errors (SSE) can be expressed as, [24]

$$\begin{aligned} \text{SSE} &= \hat{\varepsilon}^T \hat{\varepsilon} = \text{trace}(\hat{\varepsilon} \hat{\varepsilon}^T) \\ &= \varepsilon^T (I - P) \varepsilon = \text{trace}\{(I - P) \varepsilon \varepsilon^T (I - P)\} \end{aligned} \quad (10)$$

$$\text{SSE} = u^T \text{diag}(1, \dots, 1, 0, \dots, 0) u$$

$$= u_1^2 + \dots + u_{n-4}^2 \quad (11)$$

where $u = K\varepsilon$.

5.2 Distribution of the Error Sum of Squares

We next define s^2 to be the normalized error sum of squares, $s^2 = SSE/\sigma^2$; where σ is the standard deviation of the measurement errors.

If the ε_i 's are independent, normally distributed random variables, with mean 0 and standard deviation σ , i.e. $N(0, \sigma)$, then the u_i 's are also normally distributed with the same mean and standard deviation, and u_1, \dots, u_{n-4} , are independent. Thus s^2 has a chi-square distribution with $n-4$ degrees of freedom (dof).

If instead, the errors are independent, normally distributed random variables with non-zero means, $\varepsilon_i \sim N(\mu_i, \sigma)$, then s^2 has a non-central chi-square distribution with $n-4$ dof, and non-centrality parameter,

$$\lambda = \frac{\mu^T(I-P)\mu}{\sigma^2} \quad (12)$$

where

$$\mu^T = [\mu_1 \dots \mu_n]$$

The non-central chi-square probability density function is given by,

$$f(s^2) = \frac{e^{-(s^2+\lambda)/2}}{2^{(n-4)/2}} \sum_{j=0}^{\infty} \frac{(s^2)^{(n-4)/2+j-1} \lambda^j}{\Gamma(\frac{n-4}{2}+j) \cdot 2^j \cdot j!} \quad (13)$$

which reduces to the standard chi-square density function for $\lambda=0$ [25,26].

In his 1949 article in *Biometrika*, Patnaik describes several approximations to the non-central chi-square distribution which vary in their computational requirements and accuracy. For the integrity verification algorithm his "first" approximation is sufficient. In this model, the density function of the non-central chi-square variable, s^2 , with $(n-4)$ degrees of freedom and non-centrality parameter, λ , is approximated by,

$$f(s^2) \approx \rho f_{\chi^2}(s^2/\rho) \quad (14)$$

which is assumed to have a chi-square distribution with v degrees of freedom where,

$$\rho = 1 + \frac{\lambda}{(n-4) + \lambda}$$

$$v = (n-4) + \frac{\lambda^2}{(n-4) + 2\lambda}$$

Note that v , the of degrees of freedom in the approximation, is not necessarily an integer.

The corresponding probability distribution is obtained by numerical integration of the density function given by Equation 10, or interpolation in tables.

5.3 Range Residual Parameter

The range residual parameter, r , used as the test statistic for failure detection and isolation (D/I) is defined as follows,

$$r = \sqrt{\frac{\varepsilon^T \varepsilon}{n-4}} = \sqrt{\frac{s^2 \sigma^2}{n-4}} \quad (15)$$

The principal advantage of this parameter, r , as a test statistic, is that its theoretical distribution is relatively insensitive to the user/satellite geometry. As explained previously, the normalized error sum of squares, s^2 has a χ^2 distribution with $n-4$ dof when the measurement errors are normally distributed with zero-mean. Of course the actual value of r does depend on both the particular geometry and the pseudo range measurement errors. In addition, if the measurements are biased, as we have determined they indeed are, the non-centrality parameter does depend strongly on the geometry and the actual measurement biases.

Knowledge of the theoretical distribution of the range residual provides important insight into how well it will serve as test statistic. Actual implementation of the algorithm requires numerical values for detection and isolation thresholds of the test statistic. These are best determined by Monte Carlo simulation

results. Appropriate thresholds may then be selected based on requirements for probability of false alarm

and missed detection.

6.0 DETECTION / ISOLATION PROCEDURE

The D/I procedure is extremely straightforward. The 5 steps are given below, followed by a discussion of how threshold values for the test statistics can be selected.

6.1 Procedure

The user may perform integrity checking and failed satellite isolation as follows:

1. Compute the residual parameter, r , using all satellites in view (6 or more) from Equations 3-6.
2. If r is less than the detection threshold, r_D , assume that all satellites are operating properly, and the integrity check is completed. If r is greater than r_D , a failure has been detected.
3. If a failure is detected in (2) compute the residual parameters, r^1, r^2, \dots, r^n for the n subsets of $n-1$ satellites.
4. If one of the r^i 's is less than the isolation threshold, r_I , and all others are larger than r_I , the satellite omitted from the i 'th subset is the failed one. If two or more of the residual parameters are below the threshold, the failed satellite cannot be isolated.
5. If a failed satellite is detected and isolated, use the navigation solution formed by omitting the failed one. If a failed satellite is detected but cannot be isolated, use the all in view solution if necessary, but recognize that the positioning accuracy is degraded.

6.2 Threshold Selection

The most effective way to select the threshold values, r_D and r_I , is based on Monte Carlo simulation results. The difficulty in applying statistical theory to the actual selection of threshold values of the test statistic is due to the fact that the GPS pseudo range measurements actually have a slowly varying component. Measurement biases significantly affect the expected distribution; but of course the user has no way of determining their values.

A user should select detection and isolation thresholds based on his requirements for false alarm and missed detection. In this section we present extensive Monte Carlo computer simulation results which facilitate the choice of appropriate threshold values.

The simulation program assumes a 3×8 uniform satellite constellation, and a surface user with elevation mask angle of 7.5 deg. Range errors are modeled as normally distributed with $\sigma = 0.4$ m, and means ranging from -5m to +5m with uniform probability. Note that we did not model errors which may be caused by selective availability. Every 15 minutes, 100 sets of measurements are generated using this range error model. In most cases the user is located at San Francisco Airport. During the 24 hr simulations 16600 data points are generated. A number of runs performed for the Chicago area produced almost identical results.

Figure (3) shows the radial position errors vs r for nominal measurement errors at both San Francisco (SFO) and Chicago over a 24 hr period. It is clear that the residual parameter is always less than 8m, and the radial position error is less than 20m.

Figures (4-6) illustrate position errors vs r at SFO when biases of 100m, 50m, and 25m are added to one of the satellite ranges. At each time step, 100 runs are performed with each of the satellites in view in turn designated as the failed one. For example, with 6 satellites in view at time 0, 600 data points are generated; if 8 satellites are visible 15 minutes later, 800 data points are generated at that time.

Figure (7) shows the probability of false alarm and missed detection as a function of detection threshold or several satellite biases. The probability of false alarm (P_{FA}) is the likelihood that a satellite failure will be declared when all measurements are actually in the nominal range. The probability of missed detection (P_{MD}) to a given bias is the likelihood that no failure will be detected when one of the measurements actually is biased by the specified amount. For example, if the detection threshold is set at 8m, satellite biases of 100m will always be detected; 50m biases will be detected 99.9% of the time, 37.5m biases 98.5% of the time, and 25m biases 75% of the time.

If a failure is detected, the user computes the residual parameter for each subset of $n-1$ satellites. Figure (8) illustrates the position errors and residual parameters for the subset containing only nominal satellite measurements. Notice that the residual parameter for these nominal subsets still is always less than

10m. Figures (9,10) show the position error vs r for all $n-1$ subsets which contain a failed satellite with bias errors of 100m and 50m respectively. For the 100m bias error, 94.7% of the subset range residuals are greater than 10m. For the 50m bias error, 86.8% of the subset range residuals are greater than 10m.

If only one of the subset solutions has a residual parameter less than the isolation threshold it must be the one consisting of only nominal measurements. Thus the user can successfully remove the biased satellite from the measurement set. If two or more subsets are below the threshold, at least one will contain the failed satellite. The user will then realize that failure isolation is not possible. Thus the algorithm ensures that a healthy satellite will never be mistakenly removed from the navigation solution.

Table 3 summarizes the Monte Carlo simulation results when the algorithm is implemented with $r_D = 8m$, and $r_I = 10m$. The probabilities of missed detection (MD), successful detection and isolation (OK), and successful detection without isolation (NI) are given.

Table 3. Probability of Detection and Isolation with $r_D = 8m$, $r_I = 10m$			
BIAS (m)	PROBABILITY (%)		
	Missed Detection	Bias Detected & Isolated	Bias Detected NOT Isolated
100	0.0	72.2	27.8
50	0.06	50.5	49.5
37.5	1.3	34.2	64.5
25	23.2	6.4	70.4

The important question which remains is, "What impact does this process have on the user's position error?" This issue is addressed in the discussion below.

7.0 DISCUSSION

From the figures in the previous section we determined that a detection threshold value of $r_D = 8m$, would produce good results in terms of low probability of false alarm and missed detections. An isolation threshold $r_I = 10m$, produced fairly good results in terms of a user's ability to remove the biased measurement. When successful detection and isolation is accomplished, the user is 99.95% sure that his position errors will be smaller than 25m (Figure 8). When a failure is detected but cannot be isolated, the user knows that his navigation solution is not reliable but can do nothing to improve it.

The primary objective of a large group of GPS users is to minimize the error in the overall navigation solution. While in general, the presence of a biased pseudo range measurement will degrade the accuracy of the navigation solution, the user/satellite geometry may be such that even a large bias will have only a minor effect on position accuracy. In addition if the subset geometry without the failed satellite is extremely weak, positioning errors may be worse if the biased measurement is not used in forming the navigation solution.

In Figure (1) we see that, given the experimentally determined nominal error model, the magnitude of the positioning errors using all satellites in view is less than 20m. If one of the measurements has a bias of 100m, it will be detected. If the measurement bias is less than 100m there is a finite probability that it will not be detected. Figures (11-13) show the probability of incurring various radial position errors when biases of 50m, 37.5m and 25m are not detected. Results are given for detection threshold values of 4m to 12m. The sum of the probabilities indicated on each of the curves represents the probability of missed detection for the particular bias and detection threshold. Notice that although the probability of missed detection for a bias of 25m is rather large, the resulting position errors are generally within the range of nominal navigation errors.

If the satellite is isolated, and the biased pseudo range is removed from the measurement set, the resulting position errors are less than 25m, 99.95% of the time as seen in Figure (6). If the satellite cannot be isolated, the user will have to compute position from measurements to all the satellites, resulting in positioning errors of up to 310m for a 100m bias and 160m for a 50m bias as in Figures (7,8). The important

aspect of this is that he will realize that his position solution is not reliable.

Figures (14-17) compare the distribution of the positioning errors which result from an all in view solution and the integrity checking algorithm solution, in the presence of one measurement biased by 100m, 50m, 37.5m, and 25m respectively. Each chart represents the results of 6600 Monte Carlo runs. The curve marked "ALL" shows the probability density for positioning errors which would be obtained by a user who forms an all in view solution and does no integrity checking. The area under this curve is equal to 1.0.

The integrity check algorithm follows steps 1-5 described in Section 3 to detect a satellite failure and to isolate the biased measurement if possible. If no bias is detected we have a missed detection because in the simulation one of the satellites is always biased. In this case the user forms a solution based on all in view and does not know that something is amiss. If a failure is detected but cannot be isolated, the all in view solution is again used. In this situation however, the user is aware that the navigation solution is unreliable. If isolation is successful, the failed satellite is removed from the measurement set used to form the solution. The three curves labeled MD, NI, and OK show the conditional probability density functions for these three possible outcomes of the D/I algorithm. The sum of the areas under the three curves is equal to 1.0. It is clear that the algorithm has significantly reduced the likelihood of large navigation errors. In addition, when large errors are unavoidable, the user is aware that the solution is unreliable.

Lowering the user's elevation mask angle can significantly improve performance by increasing the number of satellites visible. Figure (18) illustrates the probability density of the errors for a user with a 0 degree mask angle in the presence of a 50m bias. By comparing this with the results shown in Figure (13) for a user with 7.5 degree elevation mask angle, we can quickly notice the reduction in mean error and the increase in the number of successful detections and isolations. The user can detect and isolate 78.6% of the time with a 0 degree mask, as compared to only 50.5% of the time with a 7.5 degree mask.

8.0 CONCLUSIONS

We have presented a practical new approach to GPS integrity checking. Four key subjects have been addressed: 1) motivation for integrity monitoring and possible approaches, 2) experimental determination of the measurement error structure, 3) theoretical distribution of the range residual parameter, 4) a practical means of implementing a D/I scheme based in this parameter. Each of these topics was essential in developing a sound methodology for solving the problem of GPS navigation reliability.

For users with carrier tracking C/A code receivers, such as the Trimble 4000S, the results presented here are directly applicable. The error structure of a different class of receiver will produce different results; however the framework for selecting detection and isolation thresholds and for evaluating the algorithm performance is identical to that presented here. For this reason it is important for receiver manufacturers and users to clearly understand the characteristics of the measurement errors.

The algorithm described is a snapshot approach to integrity monitoring. Under certain circumstances the performance may be improved by averaging the test statistic over several measurement times. This would be particularly useful for a code only receiver, in which the measurement errors are dominated by white noise. In the case of a carrier tracking receiver, we have found the measurement errors to be slowly varying in nature with a low level of high frequency noise. In this situation it is not likely that time averaging will cause much improvement.

The range residual based integrity verification method is reliable and easy to implement as part of a navigation software package. It will never degrade a user's solution and can frequently eliminate biased measurements, allowing a user to continue on a critical phase of his mission. When isolation is not possible a user may have to postpone navigation critical maneuvers. The only dangerous condition arises from the unlikely occurrence of a missed detection. Further studies may produce a method for predicting when missed detections are more likely, based on subset solution geometry [27]. By implementing the integrity check the user will have an accurate measure of the reliability of the navigation solution.

9. REFERENCES

1. Shively, C.A., Draft of "Comparative Analysis of Methods for GPS Integrity," Mitre Corp, October 1987.
2. Rockwell International Corp., *Navstar GPS Space Segment/Navigation User Interfaces*, ICD-GPS-200, 26 September 1984.
3. Jorgensen, Paul S., Private Conversation October 1987.

4. Braff, Ronald and c. Shively, "GPS Integrity Channel", *Navigation*, Journal of the Institute of Navigation, Vol. 32, No.4, Winter 85-86, pp. 334-350.
5. Miller, Karen J., "Independent Ground Monitor Coverage of Global Positioning System (GPS) Satellites for Use by Civil Aviation," *Proceedings of IEEE Plans*, Las Vegas, 1986.
6. Jorgensen, Paul S., "Achieving GPS Integrity and Eliminating Areas of Degraded Performance," *Navigation*, Journal of the Institute of Navigation, Vol. 34, No.4, Winter 87-88, pp. 297-306.
7. Klass, Philip J., "Industry Devising GPS Receivers with Hybrid Navigation Aids," *Aviation Week & Space Technology* 14 December 1987, p121-3.
8. Klass, Philip J., "Defense Dept. Will Seek Funds to Expand NAVSTAR Constellation," *Aviation Week & Space Technology* 5 October 1987, p30.
9. Kalafus, Rudolph M and Gerald Y. Chin, "Performance Measures of Receiver-Autonomous GPS Integrity Monitoring," *Proceedings of the ION National Technical Meeting*, Santa Barbara, 1988.
10. Lee, Young C., "Analysis of Range and Position Comparison Methods as Means to Provide GPS Integrity in the User Receiver," *Proceedings of ION 42nd Annual Meeting*, Seattle, 1986.
11. Lee, Young C., Mitre Corp Memorandum W46-M4380 25 Feb 1986.
12. Lee, Young C., Mitre Corp Memorandum W46-M4422 4 Apr 1986.
13. Brown, R. Grover and Patrick Y.C. Hwang, "GPS Failure Detection by Autonomous Means Within the Cockpit," *Proceedings of ION 42nd Annual Meeting*, Seattle, 1986.
14. Brown, R. Grover and Paul W. McBurney, "Self-Contained GPS Integrity Check Using Maximum Solution Separation as the Test Statistic," *Proceedings of the ION Satellite Division Technical Meeting*, Colorado Springs, 1987.
15. McBurney, Paul W. and R. Grover Brown, "Self-Contained GPS Failure Detection: The Kalman Filter Approach," *Proceedings of the ION Satellite Division Technical Meeting*, Colorado Springs, 1987.
16. Brown, Alison K., "Receiver Autonomous Integrity Monitoring Using a 24-Satellite GPS Constellation," *Proceedings of the ION Satellite Division Technical Meeting*, Colorado Springs, 1987.
17. Parkinson, Bradford W. and Penina Axelrad, "Simplified GPS Integrity Checking with Multiple Satellites," *Proceedings of the ION National Technical Meeting*, Dayton, 1987.
18. Parkinson, Bradford W. and Penina Axelrad, "A Basis for the Development of Operational Algorithms for Simplified GPS Integrity Checking," *Proceedings of the ION Satellite Division Technical Meeting*, Colorado Springs, 1987.
19. Parkinson, Bradford W. and Penina Axelrad, "A Practical Algorithm for Autonomous Integrity Verification Using the Pseudo Range Residual," *Proceedings of the ION National Technical Meeting*, Santa Barbara, 1988.
20. Trimble Navigation, *User's Guide to Operation of the 4000S Surveying Receiver*, 1987.
21. Geophysical Survey Inc, Preliminary Survey Report for TAU Corporation August, 1987.
22. Strang, Gilbert, *Linear Algebra and Its Applications*, Academic Press Inc, New York 1980.
23. Cook, R. Dennis and Sanford Weisberg, *Residuals and Influence in Regression*, Chapman and Hall, 1982.
24. Kshirsagar, Anant M., *A Course in Linear Models*, Marcel Dekker Inc, 1983.
25. Patnaik, P.B., "The Non-Central χ^2 - and F-Distributions and Their Applications," *Biometrika*, Vol 36, 1949.
26. Kendall, Maurice G. and Alan Stuart, *The Advanced Theory of Statistics*, Vol 2, Harner Publishing Co, 1973.
27. McBurney, Paul W., Private Conversation January, 1988.

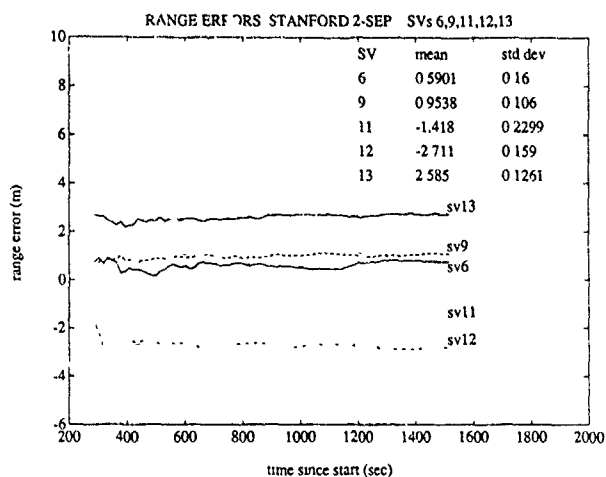


Figure 1. Typical Experimental Pseudo Range Error Trace - Nominal Aided Mode (2 Sept, 1987).

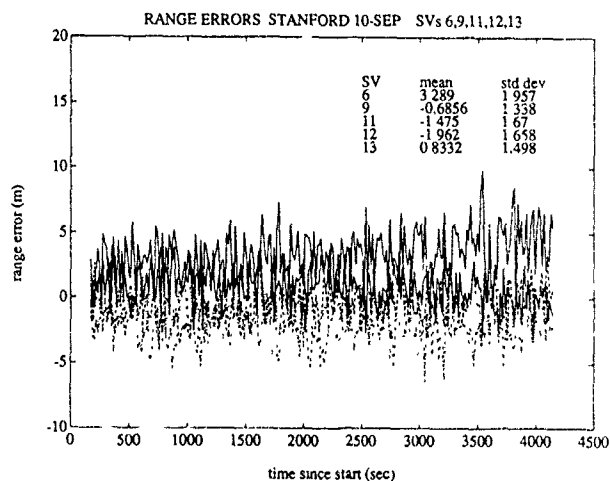


Figure 2. Typical Experimental Pseudo Range Error Trace - Nominal Unaided Mode (10 Sept, 1987).

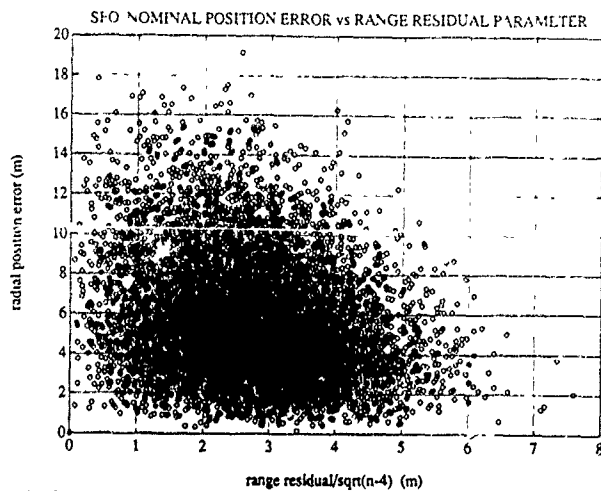


Figure 3a. Nominal Position Error vs Residual Parameter for San Francisco - All in View Solution.

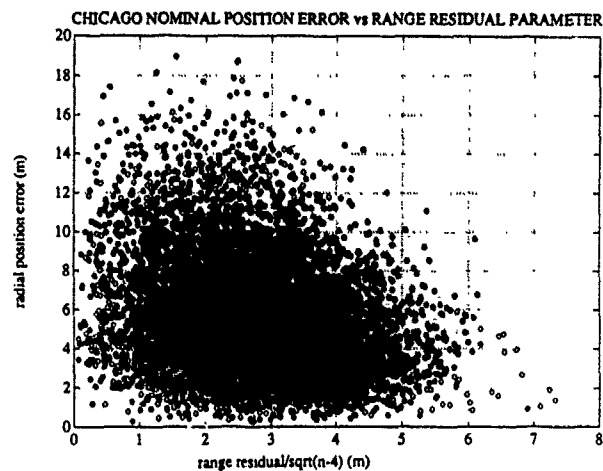


Figure 3b. Nominal Position Error vs Residual Parameter for Chicago - All in View Solution.

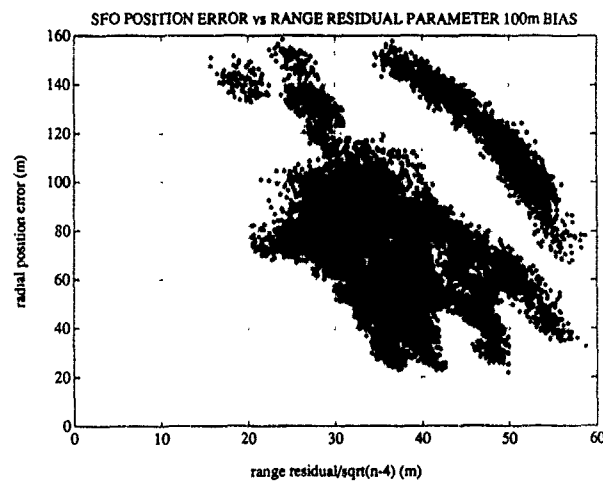


Figure 4. Position Error vs Residual Parameter with 100m Biased Measurement. 16600 data points are shown for a 24hr time period. Every 15min 100 runs are done w/each satellite in view designated as failed.

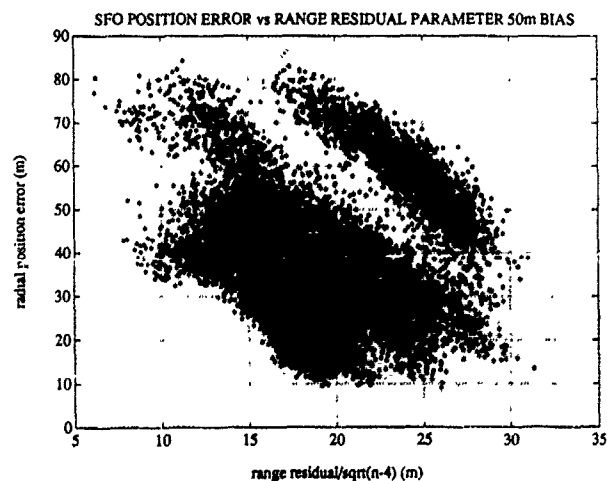


Figure 5. Position Error vs Residual Parameter with 50m Biased Measurement.

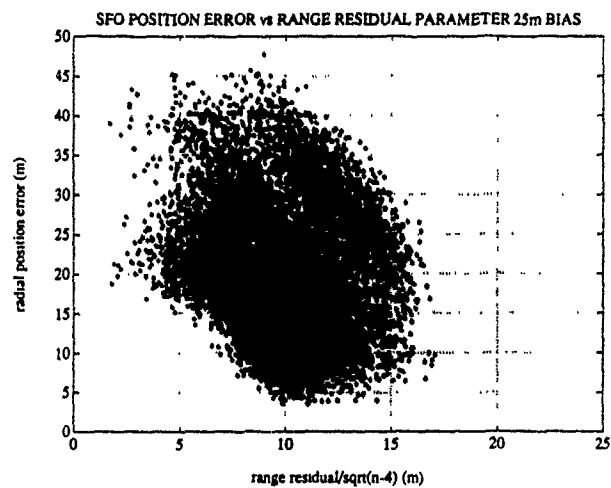


Figure 6. Position Error vs Residual Parameter with 25m Bias.

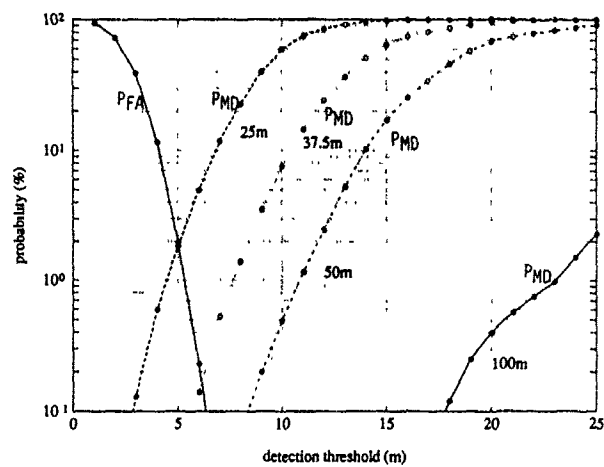


Figure 7. Probability of False Alarm & Missed Detection vs Detection Threshold for Biases: 100,50,37.5,25 m.

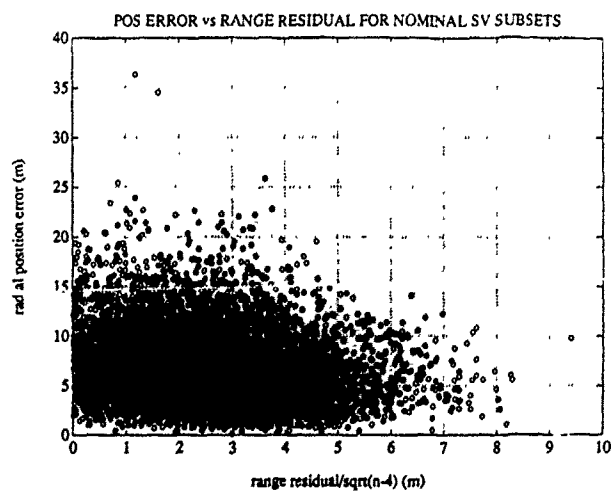


Figure 8. Position Error vs Residual Parameter for Satellite Subsets Which Do NOT Include the Biased Satellite.

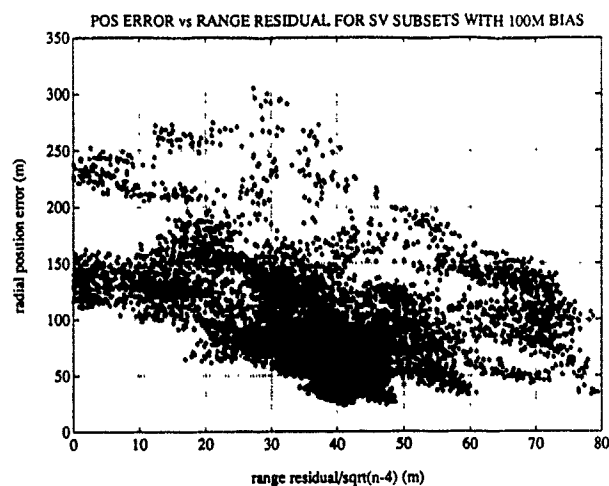


Figure 9. Position Error vs Residual Parameter for Subsets Including Satellite with 100m Bias.

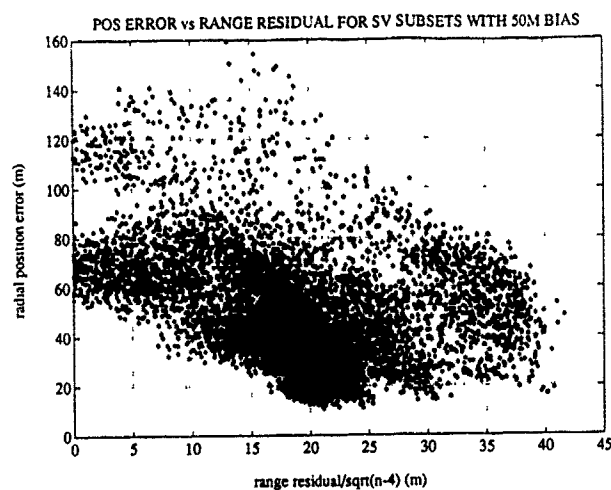


Figure 10. Position Error vs Residual Parameter for Subsets Including Satellite with 50m Bias.

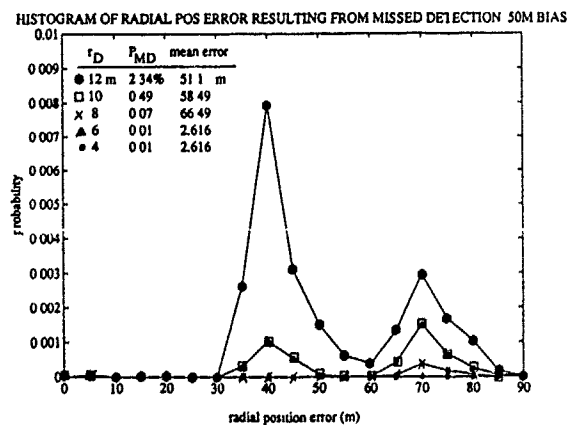
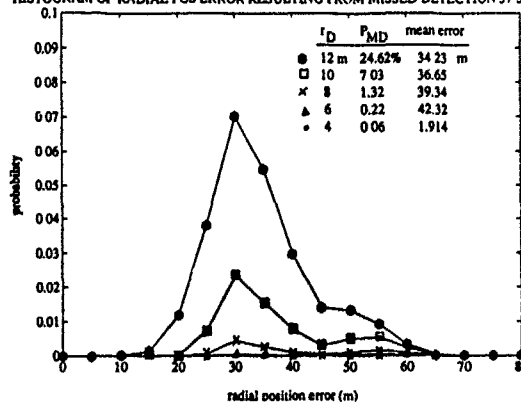


Figure 11. Distribution of Radial Position Errors When a 50m Bias is NOT Detected. Detection Thresholds $r_D = 4, 6, 8, 10, 12$ m.

HISTOGRAM OF RADIAL POS ERROR RESULTING FROM MISSED DETECTION 37.5M BIAS

Figure 12. Distribution of Radial Position Errors When a 37.5m Bias is NOT Detected. Detection Thresholds $r_D = 4, 6, 8, 10, 12$ m.

HISTOGRAM OF RADIAL POS ERROR RESULTING FROM MISSED DETECTION 25M BIAS

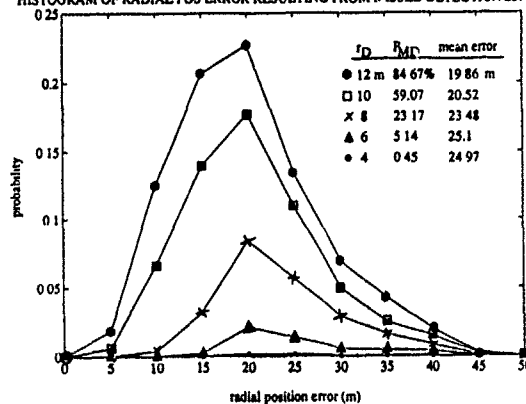
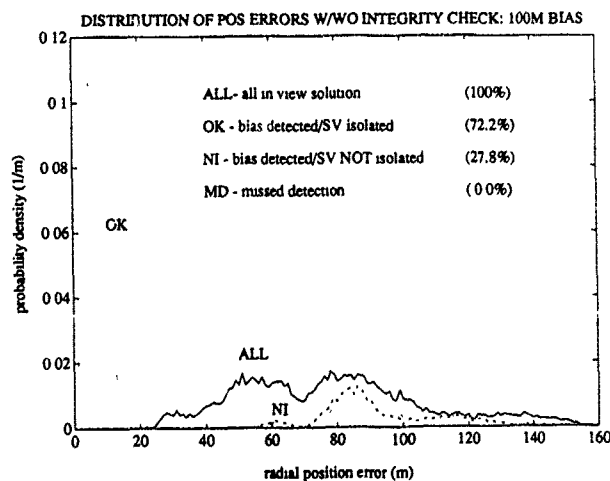
Figure 13. Distribution of Radial Position Errors When a 25m Bias is NOT Detected. Detection Thresholds $r_D = 4, 6, 8, 10, 12$ m.

Figure 14. Probability Density Function for Radial Position Errors. All in View Solution, and Conditional Probability Density Functions for Radial Position Errors from Possible Outcomes of Integrity Checking Algorithm. One Satellite has a 100m Bias.

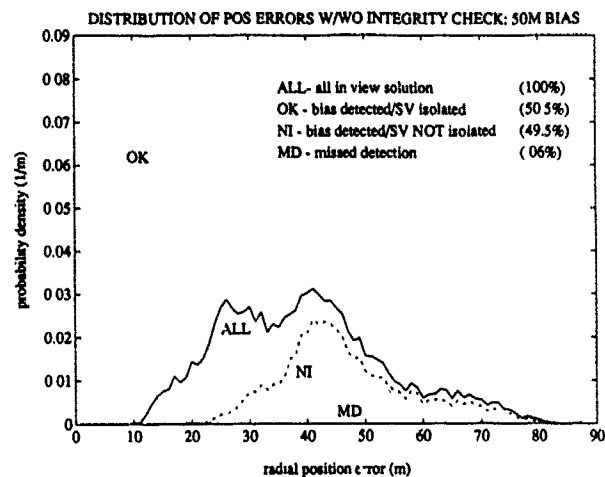


Figure 15. Probability Density Functions for Radial Position Errors. One Satellite has a 50m Bias.

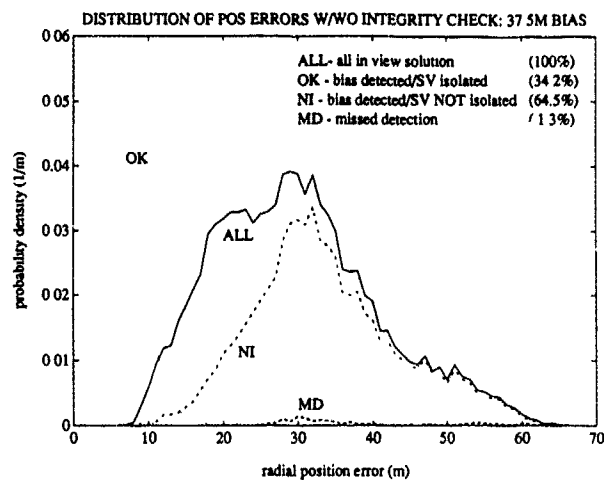


Figure 16. Probability Density Functions for Radial Position Errors. One Satellite has a 37.5m Bias.

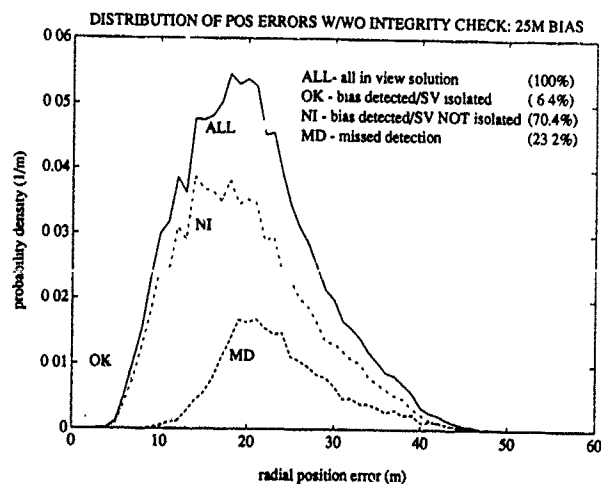


Figure 17. Probability Density Functions for Radial Position Errors. One Satellite has a 25m Bias.

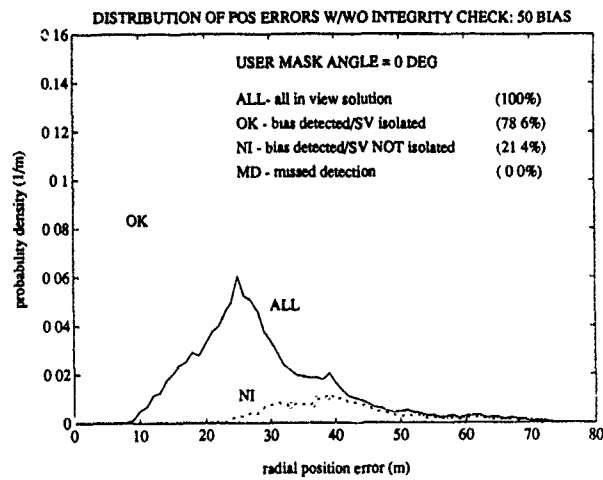


Figure 18. Probability Density Functions for Radial Position Errors. One Satellite has a 50m Bias. User Elevation Mask Angle is 0.0 Degrees. 9700 Data Points.

MODULAR DIGITAL GPS RECEIVERS

by

J.Scott Graham, Staff Engineer
 Peter C.Ould, Principal Engineer
 Robert J. Van Wechel, Senior Chief Scientist
 Interstate Electronics Corporation
 1001 East Ball Road
 Anaheim, California 92803
 United States

ABSTRACT

This paper describes the modular digital approach to GPS receiver design being implemented at Interstate Electronics Corporation (IEC) for various range instrumentation and military P-code applications.

The receiver consists of a preamplifier, RF downconverter that converts the GPS signal spectrum to baseband and digitizes it, single or multiple digital tracking processors for carrier and code tracking, a receiver control and navigation processor, and various types of flexible modular interface (FMI) boards.

This receiver is IEC's third generation P-code design and has been achieved by a graceful miniaturization process that utilizes currently available low risk technology to reduce size, power consumption, and cost. Recent anti-jamming test results are presented.

1.0 INTRODUCTION

IEC is developing a new (L1/L2/P-code) GPS receiver that is ideally suited to navigation and positioning applications in platforms such as missiles, remotely piloted vehicles, aircraft, aircraft pods, and ships. This receiver is being developed as part of IEC's Independent Research and Development (IRAD) program.

The receiver concept is highly modular and includes provisions for low, medium, and high dynamics operations as shown in figure 1. Low dynamics applications (in the 1g range) can be handled by the single channel receiver unit, which is a 5-inch diameter by 11.5-inch long module, as shown in figure 2. For these dynamics, the receiver operates as a single channel, fast multiplex receiver. The antenna and preamplifier unit is a separate package. For medium dynamics (in the 1 to 4g range), a small, low cost, inertial reference unit can be added to provide an inertially-aided receiver that will track these dynamics. High dynamics applications require the use of the 5-channel receiver unit which can be packaged in a 5-inch diameter by 19.5-inch long unit for aircraft pod applications, or in a 1/2 ATR short package. For these dynamics, the receiver operates as a parallel continuous processing receiver. All of these receivers include provisions for differential processing.

Special features are included to enhance acquisition speed and anti-jamming. These include a VLSI digital parallel correlator to compare the signal with the local reference code, significantly speeding up acquisition time and time-to-first-fix. Special anti-jamming features have been included that particularly improve anti-jamming performance in the presence of CW jamming, which has been the worst case jamming threat in other available GPS receivers.

Provision is included in the receiver design to handle the selective availability and anti-spoofing functions that deny high accuracy to unauthorized users in times of a national emergency.

The control of the receiver can be either by a control/display unit in applications that require an operator interface, or by a computer interface in computer-controlled applications. Both control methods use flexible modular interface (FMI) boards.

2.0 ARCHITECTURE

A block diagram of the receiver is shown in figure 3. The units shown are the preamplifier module and either the single channel receiver unit or the five channel receiver unit. The preamplifier module shown operates from two switched (multiplexed) antennas, such as top and bottom aircraft or pod antennas. Other preamplifier types are available to employ a single antenna. The family of modules for the receiver is shown in figure 4.

Referring again to figure 3, the input L-band signals from the preamplifier are downconverted to in-phase and quadrature baseband signals in the RF converter, then digitized in the adaptive analog-to-digital converter (ADC). These digitized signals are then processed in the tracking processor, which contains code and carrier tracking loops to track the GPS signals. For single channel receivers, the tracking processor operates in a fast multiplex mode, dwelling on each satellite for a few milliseconds before going on to process the next satellite signal. A similar multiplex approach is used

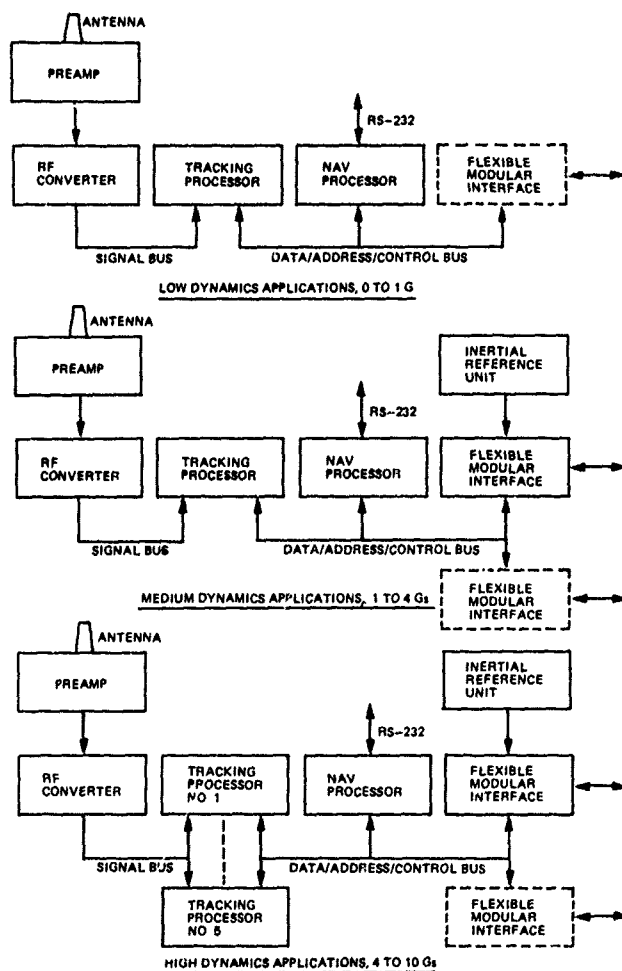


Figure 1. Modularity for Various Applications

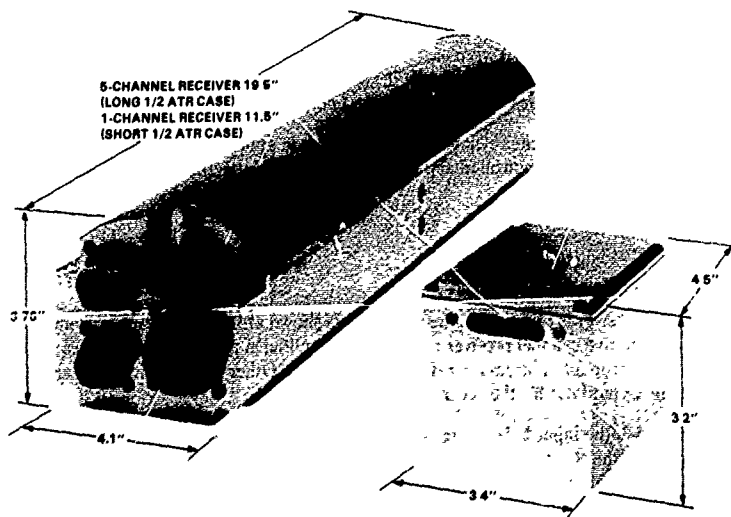


Figure 2. Five-Channel Receiver Unit of Modular GPS Receiver and Battery Power Assembly

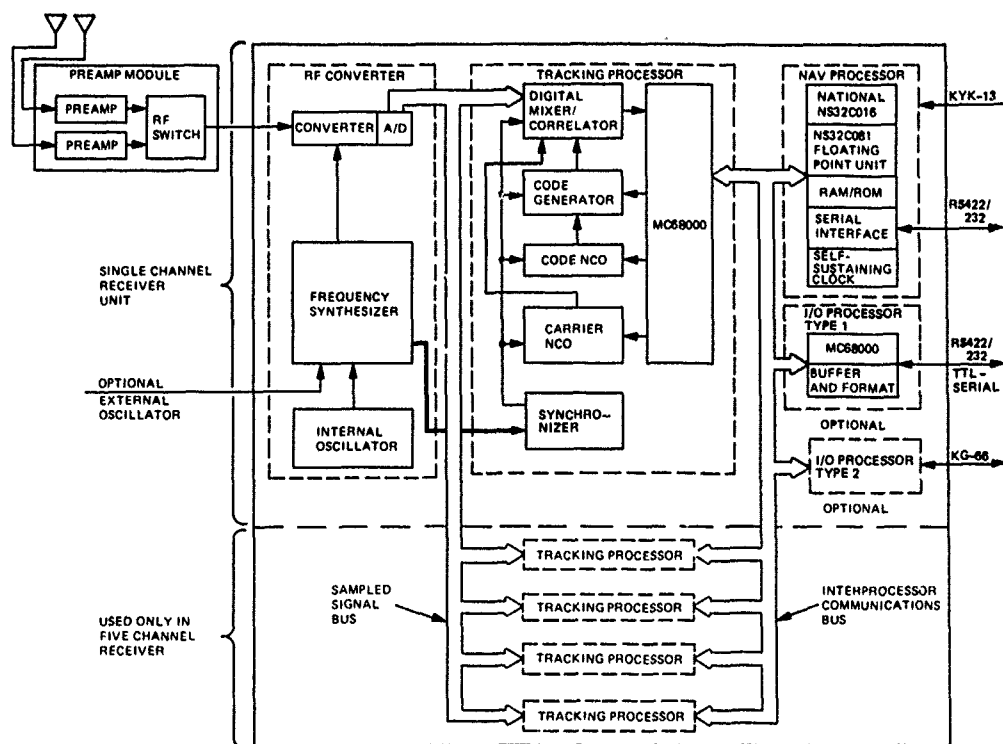


Figure 3. Block Diagram of IEC Advanced GPS Receiver

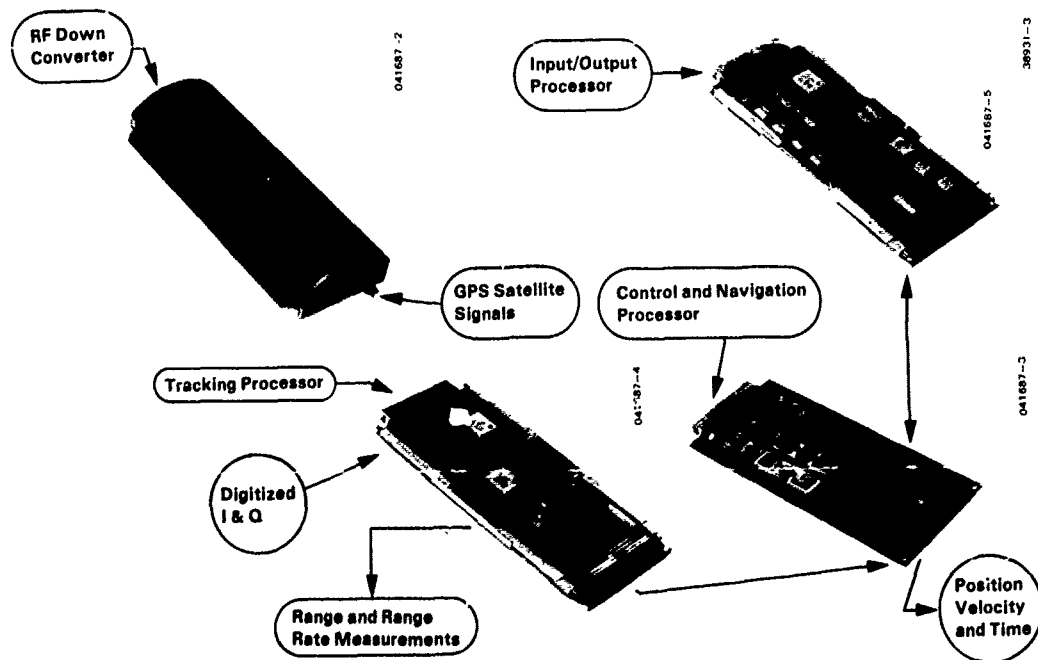


Figure 4. Family of Modules for the IEC Advanced GPS Receiver

for two channel receivers, except that each tracking processor then only needs to multiplex half as many satellites, which reduces the multiplexing loss. This fast multiplexing approach provides a capability for high dynamic operation with reduced hardware. For five or six channel receivers, there is one tracking processor for each satellite signal. Fast multiplexing is used in the five or six channel receiver for multiplexing between L1 and L2 frequencies, and between top and bottom aircraft or pod antennas.

The final conversion to remove doppler is performed digitally by each individual tracker, eliminating the need for multiple IF channels.

The tracking processor code and carrier tracking loops are closed via the CMOS 68000 microprocessor. The carrier loop feeds back through the carrier numerically-controlled oscillator (NCO) whose digital output is used to remove the doppler from each individual satellite signal. The code tracking loop feeds back through the code NCO and code generator. The 68000 is on a common interprocessor communications bus with the navigation processor and FMI processors, allowing these processors to access memory for control and data. The output measurements from the tracking processor are the pseudorange and delta pseudorange measurements, which are accessed by the navigation processor over the bus and are used as inputs to the navigation filter.

The navigation processor uses a CMOS National Semiconductor NS32C016 microprocessor, which was selected for high throughput capability in floating point computations when used with its floating point coprocessor. This processor is used for management processing and for the GPS navigation filter. The battery powered self-sustaining clock is also contained in the navigation processor.

The FMI boards all employ a CMOS 68000 processor with direct memory access to allow rapid communication of I/O data over the interprocessor bus.

Special function boards can be added. As an example, a 4-Mbyte RAM board is used as a solid state flight recorder in some applications instead of using a separate external recorder.

The fixed frequency downconversion, the local carrier and code tracking for multiple tracking processors, and the separation of navigation, tracking, and input/output processors are cornerstones of a highly modular receiver architecture. The allocation of software tasks to each distributed processor allows for software modularity. The single RF downconverter can drive from one to N tracking processors with no loss of performance for any channel. This allows for appreciable power savings over the alternate approach of several RF downconverter channels. Hardware fault isolation is simple because modularity allows swapping of tracking processor boards. Shifting of signal processing hardware from analog/RF circuitry to VLSI CMOS digital logic lowers parts count, improves reliability, lowers power dissipation, reduces component drift, and simplifies calibration. Technology insertion of higher density VLSI components can have minimal impact on receiver structure.

3.0 DETAILED DESIGN

3.1 Antenna

A wide variety of antennas can be used with the receiver. For high dynamic aircraft and pods, a top- and bottom-mounted antenna pair is used. In other applications, a single antenna is sufficient.

For top- and bottom-mounted antennas, a pair of preamplifiers feeding a solid-state switch are used. Preamplifier gain is 30 dB, and an overall noise figure of 3.1 dB is achieved. The solid-state antenna switch allows switching antennas in 200 ns, so negligible loss is taken in fast multiplexing antennas with a typical dwell time on each antenna of 20 ms or less.

3.2 RF Converter

The RF converter block diagram is shown in figure 5. A diplexer is used at the input to separate the L1 and L2 signals, which are then multiplexed by a solid-state switch into the first mixer. The first conversion reference frequency is fixed at 1401.51 MHz that converts both L1 and L2 inputs without switching reference frequencies. Following downconversion, amplification and bandpass filtering, the signal is split into C/A- and P-code channels for the final conversion to baseband. Following this conversion, the individual outputs are digitized to two-bit words for input to the tracking processor. There is no doppler removal or code stripping done in the RF converter. This allows the RF converter to drive any number of tracking processors, and contributes greatly to the modularity of our approach.

The analog-to-digital converter (ADC) is a special custom-integrated circuit that was designed to provide high immunity to non-Gaussian interference and jamming. It is particularly effective against CW, swept CW, and pulsed CW due to the highly non-Gaussian amplitude distribution of these types of interference.

The adaptive threshold ADC separately quantizes both the I and Q channel received signal components prior to baseband correlation as shown in figure 5. Two bits (sign and magnitude) are retained as shown in figure 6.



Following the selection of C/A- or P-code sampled input, the doppler is removed in a read-only memory (ROM) that is programmed to perform a complex multiplication that rotates the phase of the input signal by the residual doppler. The two-bit signal is then applied to the multi-tap correlator that correlates it with either the C/A- or P-code reference. The correlator outputs are input to the 68000 processor for use in both acquisition and track modes.

3.3 Tracking Processor

The tracking processor implements an all-digital baseband processing design that includes a 12-chip wide acquisition and track correlator for improved acquisition speed and dynamic tracking.^(3,4) This board tracks both code and carrier in the signal. Figure 9 shows both a simplified functional operation diagram and the gate array partitioning employed.

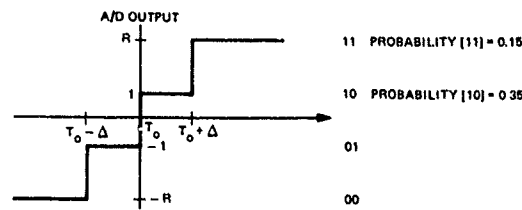


Figure 7. 2-Bit Threshold

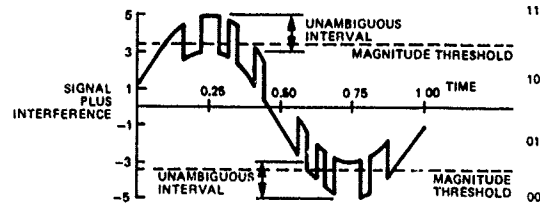


Figure 8. Output Presentation (noise not shown)

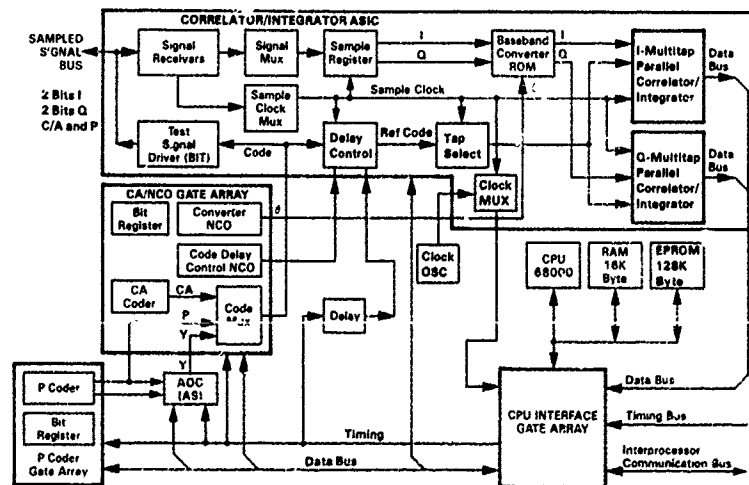
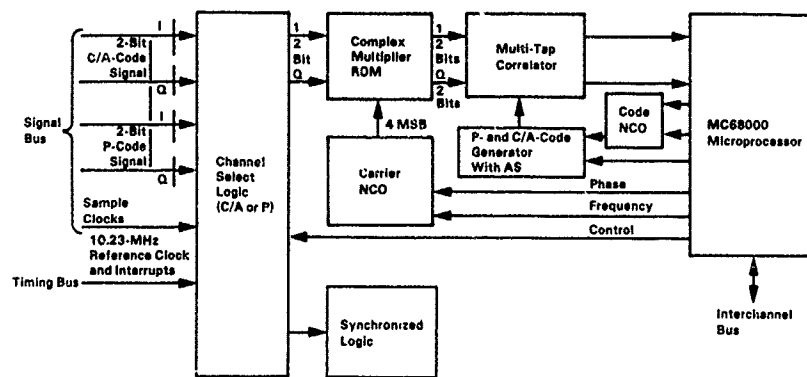


Figure 9. Tracking Processor

The code generators are designed for fast multiplexing in that they can be set to any desired chip under control of the 68000. The P-coder is contained on one CMOS gate array, and the C/A-coder with two numerically-controlled oscillators is contained on another. The anti-spoofing function is also included for operation with the P-code generator.

3.3.1 Digital Doppler Removal

The tracking processor (figure 9) performs digital doppler removal by implementing a carrier frequency lock loop (FLL), a carrier phase lock loop (PLL), or a Costas loop. The microprocessor program performs feedback computations that close the loop between the error measurement hardware (digital correlator/integrators) and the signal reference generation hardware (P-coder, C/A-coder, NCOs, code delay, and baseband converter).

The NCO consists of a 24-bit accumulator that is caused to overflow periodically at the desired output frequency. The converter (carrier) NCO is initialized with the estimated carrier doppler frequency and phase. Several of the most significant bits of the instantaneous phase output are applied as the rotation angle for the baseband converter. The function of the baseband converter (BBC), shown in figure 10, is to perform a phase rotation of the input signal. The BBC is actually implemented as a read-only memory lookup table. The I and Q inputs are the real and imaginary components, respectively, of the complex input signal, and similarly for the outputs.

The 24-bit accumulator width of the converter NCO was determined to be adequate to keep phase jitter contribution below the system phase noise level. The limited number of NCO bits used for the phase rotation angle does not cause a coarse phase resolution; the full precision phase estimate is maintained in the software digital loop filter. The negative signal-to-noise level of the BBC inputs and outputs, and the integration time averaging over carrier phase cycles allows the use of the full phase resolution of the NCO.

3.3.2 Code Delay Approach

Reference code time alignment is performed by code clock manipulation and using a delay lock loop implemented in software. The original code doppler removal technique⁽³⁾ used an NCO clocked at 8.42 MHz to generate a 1.023-MHz (plus doppler) frequency C/A-code clock. Application of this technique to P-code was not pursued since the capabilities of VLSI CMOS logic would not be sufficient. In the present design, all reference code generators are clocked by the 10.23-MHz reference clock and the code outputs are delayed by fractions of a code chip. Code doppler is simulated by stepping the code delay over time. The code generators generate integer P-chips of delay and the Code Delay NCO controls fractional P-chips of delay. This approach uses an NCO clock frequency less than 1 MHz for doppler control of both C/A- and P-code. The setting of code delay rather than frequency avoids accumulation of numerical integration error. Additional benefits include low interchannel delay bias and compatibility with fast multiplexed tracking. Reference code initialization for any satellite and any code value may be set to very fine delay resolution within four milliseconds. Although code initialization computations can use a large fraction of one millisecond, the synchronized loading of the working registers occurs with no time lag penalty.

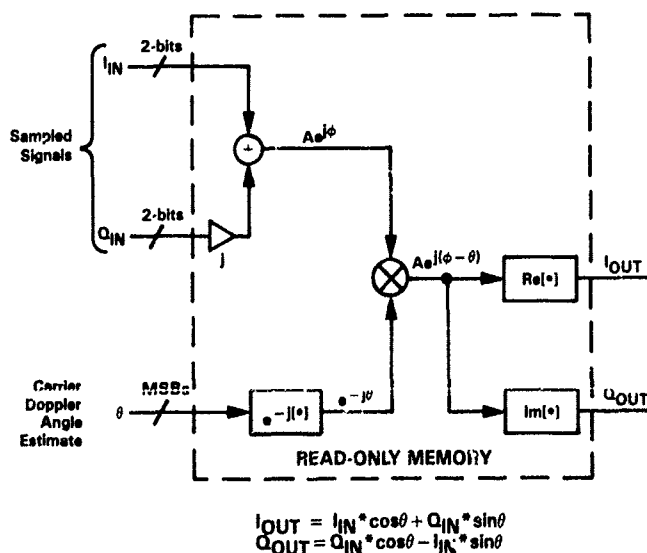


Figure 10. Digital Baseband Converter

3.3.3 Multi-tap Correlator Integrator

The Correlator/Integrator ASIC is predominantly a set of 16-bit up-down counters. When sampled signals are correlated against reference code, the counters count up or down to indicate a match or mismatch, respectively. At the end of the correlation time interval, the counts are stored in holding registers accessible over the microprocessor bus. This integrate and dump function is the optimal filter for detection of pseudorandom noise code buried in Gaussian noise.

The correlator control logic was made flexible to allow for different capabilities in different modes of operation. For example, one, two, or three I/O channels may be configured to achieve from two to eight levels of quantization for the BBC outputs. Figure 11 indicates how the channel outputs may be combined to achieve eight-level, four-level, and two-level operation. The weighting factors for this operation may be manipulated under software control to get the best processing gain from the adaptive ADC.

The multi-tap correlator integrator is possible as a consequence of the digital sampled signal approach. There are approximately two samples per code chip. Reference code sampled into an N-bit shift register gives N delayed reference code samples each clock cycle, spaced one-half code chip apart. The correlator control logic can then correlate the input sampled signal against all N reference codes in parallel. Figure 12 indicates that eleven code chips may be searched in parallel during one correlation interval. This permits a sharp reduction in acquisition time over systems that do not use multi-tap correlation. The large code chip delay range also makes possible an extended range delay discriminator. In high dynamics and low signal to noise applications, the delay lock loop using this detector will be able to "hang on" better and have lower loss of lock levels than a standard Early Minus Late discriminator.

3.3.4 Sample Frequency Considerations

The relationship of two samples per code chip is advantageous to the multi-tap correlation, but four, six, or eight samples per chip might also be used. Calculations and measurements show that, with sufficient bandlimiting of the presampled signal, there is negligible gain to be had by sampling at greater than twice the code chip frequency. Lower clock frequencies also reduce system noise levels and reduce power dissipation in the VLSI CMOS circuits.

The sample frequencies used, 20.5 MHz and 2.05 MHz for P-code and C/A-code, respectively, are asynchronous relative to the code frequencies. If the sample rate is exactly twice the pseudorandom code rate, the auto-correlation function loses its characteristic shape and becomes a staircase function.

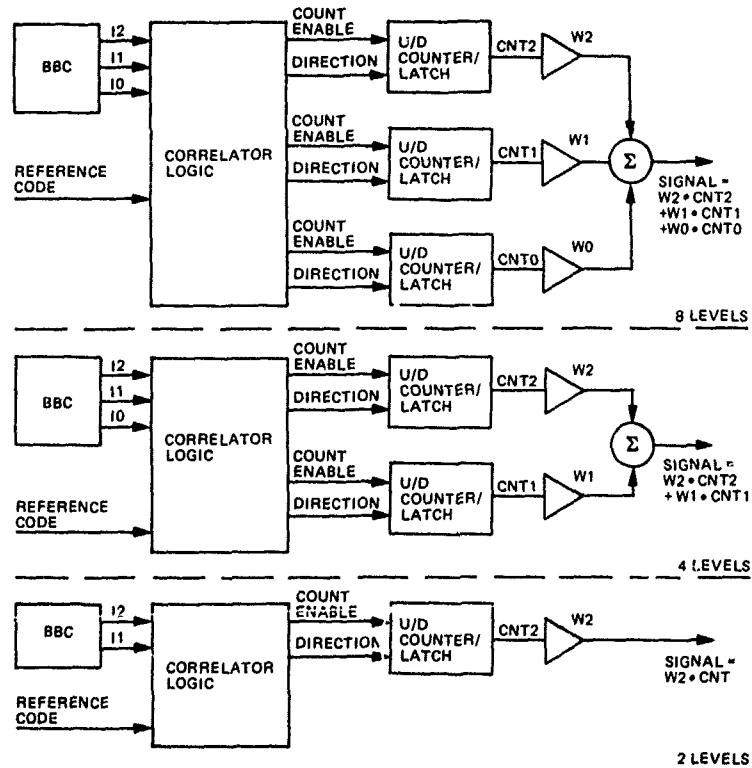


Figure 11. N-Level Correlator

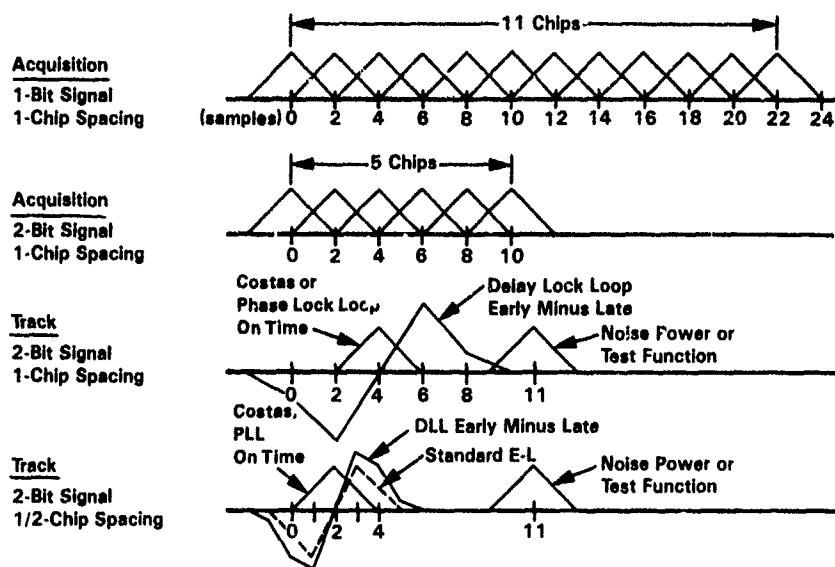


Figure 12. Correlator Tap Select

3.4 Navigation Processor

The navigation processor uses the National CMOS NS32C016 microprocessor and its associated floating point unit that provides the high floating point throughput required for high dynamics applications. This processor serves as the inter-processor communications bus controller, management processor, and navigation filter processor. The navigation processor has complete access to the memory space within the tracking processors and FMI processors.

A block diagram of the navigation processor is shown in figure 13. The following memory is contained on the board:

- 256K byte CMOS UV EPROM instruction store
- 128K byte CMOS RAM for data storage
- 16K byte battery-backed CMOS RAM for storage of critical data
- 16K byte EEPROM for both the almanac and a bootstrap used for field reprogramming of onboard EPROM

A low power battery-backed self-sustaining clock is included on the navigation processor board to maintain time greater than 30 days without prime power.

3.5 Flexible Modular Interface (FMI)

The bus-oriented architecture and the use of direct memory access in the FMI cards allows each FMI card to access all receiver data, supporting the multiple use I/O port concept and easing system integration.

The FMI is under the direction of the navigation processor, but operates autonomously up to the message level, controlling the various I/O transfers concurrently. In addition to performing the communications function, the FMI boards also provide message data converting, rescaling, and reformatting capability. The message data is originally developed by, or ultimately destined for, the navigation processor, which sees the FMI as a resource residing in its memory space. The FMI is specifically designed to minimize the navigation processor workload.

Figure 14 shows the block diagram of the FMI boards. All of the FMI boards are designed around a common structure that minimizes development and enables use of much common hardware and software.

Variations between FMI types are due to the peculiar interface requirements of each, resulting in specialized hardware and software drives only in this area.

The Motorola 68000 processor is used in all FMI boards. A key feature of the design is the direct memory access (DMA) structure, in which the separate port driver circuits have rapid access to data storage memory, transparent to the microprocessor. Thus, the 68000 is not in the loop for each data transfer, but serves as the off-line orchestrator of board activity.

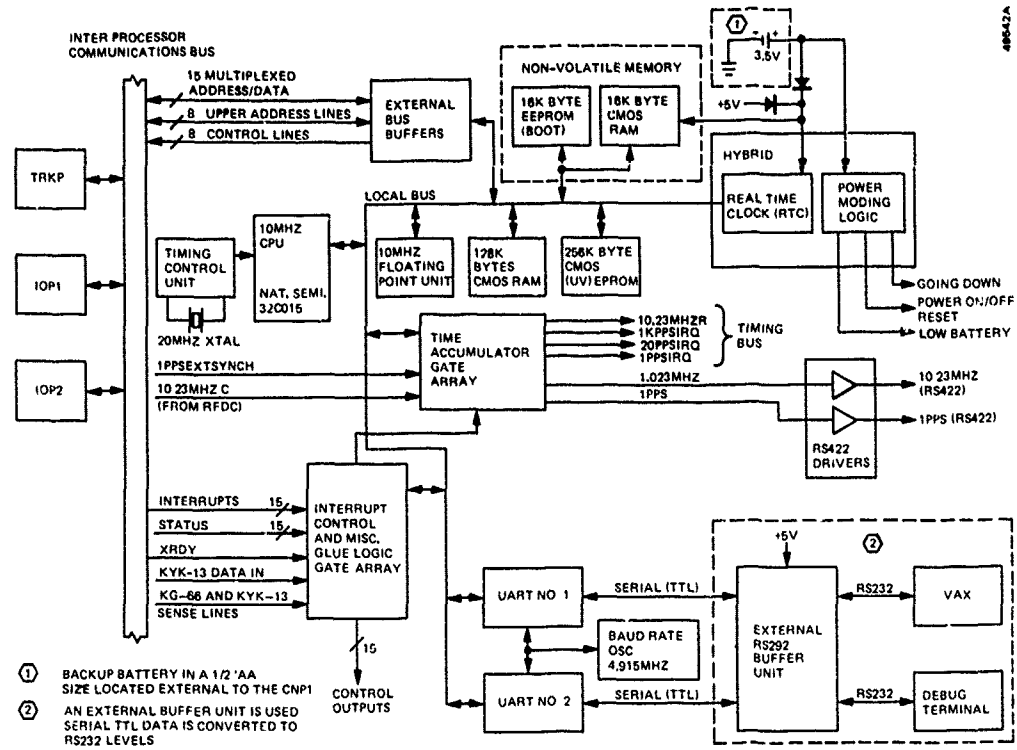


Figure 13. Navigation Processor Block Diagram

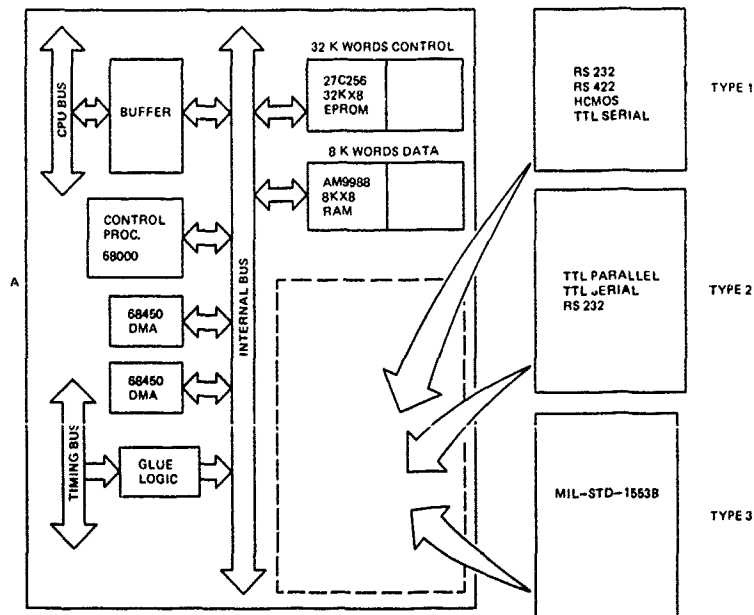


Figure 14. Flexible Modular Interface (FMI) Block Diagram

3.9 ASIC Development

The first printed circuit version of the tracking processor required roughly 300 dual inline type integrated circuits. Size constraints on the P-code Engineering Development Model unit made some form of specialized digital integrated circuit mandatory. The choices at that time were full custom in either Bipolar TTL, ECL, or CMOS and CMOS gate arrays. The CMOS semi-custom gate arrays at the upper edge of the state-of-the-art satisfied the logic speed requirements (21 MHz), and their inherent low power and relatively low cost made them an ideal choice for the ongoing development. The NCO and the first generation correlator/integrator gate arrays were developed for this project. The effort indicated clearly that the technology was well suited to the task and that it was realistic to plan on having gate array devices "right the first time".

The P-Code Reference Receiver system required a further reduction in size. CMOS gate array technology was a clear cut choice in terms of cost, power, schedule, and range of available vendors. At this point in the development, where the system became primarily a collection of semi-custom devices, the architecture and functional partitioning became critical factors. The design goal was to partition and mechanize the functions to be standalone building blocks so that various system configurations could be realized with a minimum of redesign. Each gate array would be designed to function as a peripheral to the Motorola 68000 microprocessor and incorporate built-in-test functions. Figure 16 indicates the path of gate array development during development of the 6-inch by 9-inch Tracking Processor printed circuit board.

Careful attention to design detail and the effective use of computer aided design and simulation tools, produced a set of gate array devices, all "right the first time". The integrated circuits for use in the Advanced Development Receiver are completed. With the exception of the extended range correlator/integrator, these devices are primarily support for the navigation and interface functions of the receiver.

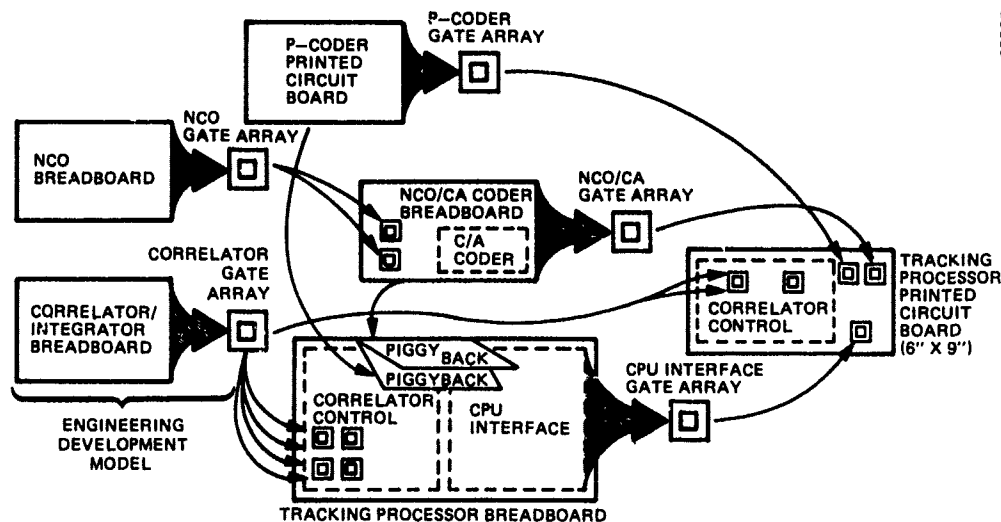


Figure 16. Reference Receiver Tracking Processor Gate Array Development

4.0 ACKNOWLEDGEMENT

The authors would like to thank the hundreds of engineers, managers, and other supporting personnel at Interstate Electronics Corporation and at our customer organizations that have contributed to this development. Without a significant team effort, none of this work would have been accomplished.

REFERENCES

1. Amoroso, "Adaptive A/D Converter to Suppress CW Interference in DSPN Spread-Spectrum Communications," IEEE Trans. on Communications, Vol. COM-31, No. 10, October 1983, pp. 1117-1123.
2. Amoroso, "Performance of the Adaptive A/D Converter in Combined CW and Gaussian Interference," 1984 IEEE Military Communications Conference, MILCOM '84, 21-24 October 1984.
3. P. C. Ould and R. J. Van Wechel, "All-Digital GPS Receiver Mechanization," Navigation: Journal of the Institute of Navigation, Vol. 29, No. 3, Fall 1981, pp. 178-188.
4. J. S. Graham, P. C. Ould, and R. J. Van Wechel, "All-Digital GPS Receiver Mechanization — Six Years Later," presented to the Institute of Navigation National Meeting, 20-23 January 1987.

INTEGRATION OF GPS/INS WITH PARTITIONED FILTERS

John W. Diesel
 Diesel Computing Systems, Inc.
 5177 Alhama Drive
 Woodland Hills, CA 91364
 USA

SUMMARY

The optimal integration of a strapdown INS with a GPS in a maneuvering vehicle is considered, assuming the GPS will eventually be jammed. To achieve optimal performance after jamming, many inertial instrument errors should be calibrated by the GPS. Instead of using a single, high order Kalman filter a more effective solution is obtained by using several partitioned, low order filters. This solution is also applicable to other GPS/INS integration problems.

The observation information used in these filters comes from the phase-locked GPS carrier tracking loops. Inertial rate-aiding of these loops to delay loss of lock due to jamming is investigated. This is a more difficult problem than early investigators anticipated. It is solved by partitioning each tracking loop into a high bandwidth extrapolator for upgrading the inertial aiding information and an aided low bandwidth tracking loop for tracking the GPS phase.

To gain insight into these partitioned filters, a general class of Kalman filtering problems is investigated. In most cases, an analytic solution to both the transient and steady-state Kalman filter gains is obtained. These solutions can be implemented directly in the partitioned filters, rather than using real-time Kalman filters.

LIST OF SYMBOLS

(Does not include symbols defined by the tables or figures)

A_x, A_y, A_z	Acceleration Components (ft/sec ²) (also specific force coefficients in error model)
C_B^L, C_E^R, C_R^L	Direction cosine matrices from subscript system to superscript system
$\delta C_x, \delta C_y, \delta C_z$	Coriolis errors
D_x, D_y, D_z	Accelerometer errors in body axes (ft/sec ²)
$\Delta R, \Delta P$	GPS delta range and corresponding delta position
$\delta()$	Differential error in (variable)
$\Delta()$	Difference in (variable) over finite time
E_x, E_y, E_z	Gyro angular rate errors in body axes (radians/sec)
ϵ	Error in control loop
e_i, e_{ij}	Unit vector along line-of-sight to satellite i , and its j th component ($j=X,Y,Z$)
e^{-ST}	Laplace transform notation for delay of T seconds
G, \bar{G}	Gain ratio, normalized gain ratio
H	Observation matrix
I	Identity matrix
K	Optimal gains
N	Upper limit on index
P_x, P_y, P_z	Position components (feet)
P	Position or phase
P_a	Antenna position
P (matrix)	Covariance matrix
PSD	Power spectral density

ϕ_X, ϕ_Y, ϕ_Z	Misalignments between computed reference axes and true reference axes (radians)
ψ	Heading relative to reference axes
q_B^L	Quaternion from body axes to level axes
Q	Power spectral density of process noise
R	Power spectral density of observation noise
$\frac{1}{R}$	Reciprocal of earth radius
t	Time in seconds
t_k, t_n	Time at sampling instants
τ	Time constant
T	Time interval
V_X, V_Y, V_Z	Velocity components (ft/sec)
v	Observation noise
w	Process noise
W_X, W_Y, W_Z	Total spatial rate of reference coordinate system (radians/sec)
X	Error state variable
Z	Observation

Superscripts

B	B coordinates
L	L reference coordinates
R	No. referenced coordinates
E	Earth-centered-earth-fixed coordinates
($\hat{\cdot}$)	Estimated (variable)
($\dot{\cdot}$)	Time derivative of (variable)

1. INTRODUCTION

Many papers have been published on integration of GPS with INS (Ref. 1-4). When GPS is available, the integrated performance is determined almost entirely by the GPS. That is, the position and velocity are determined by the GPS code and carrier tracking loops, respectively, with little help from the INS except possibly during turns. The real advantage of the INS is that it cannot be jammed. Therefore, the GPS and INS should be integrated to optimize the INS performance after the GPS is jammed.

This problem was investigated in Reference 3 for a hypothetical weapon delivery example. It was assumed that a missile guidance system using GPS and inertial guidance was jammed several minutes before reaching the target. The missile was then guided by the inertial system until the target area was reached. It was shown that the position accuracy of the GPS could be maintained for 5 minutes after jamming, provided the inertial system was properly aligned and calibrated by the GPS before jamming occurred. Also, it was shown that the accuracy deteriorated rapidly 10 minutes after jamming because of gyro noise, even with a well calibrated system. Therefore, carrier loop rate-aiding was investigated as a means of delaying the time at which jamming occurred.

The missile guidance example of Reference 3 led to many interesting analysis, design, and synthesis problems which will be examined here. In particular, proper calibration of the inertial system led to a filter with up to 40 error states. The conventional approach to such problems is to use a single, large Kalman filter. Since the computer burden grows roughly as the cube of the number of states, this could become a problem even with newer, faster computers. It is therefore shown in Sections 2, 3, and 4 that the problem can be solved more efficiently and more accurately by partitioning the state variables into several smaller groups, and using separate low order filters for each.

The problem of rate-aiding the carrier tracking loop is investigated in Section 5. This leads to a problem in optimal estimation and prediction which is solved by partitioning the solution into a third order high bandwidth filter in parallel with a low bandwidth third order tracking loop. Finally, insight into the design of these partitioned filters and those in Sections 2, 3, and 4 is obtained in Section 6. This is done by solving a general class of filtering problems. The solution is determined analytically in most cases, so that the effects of changing assumptions and modeling errors can be evaluated in advance of the application.

2. STRAPDOWN INERTIAL SYSTEM ERROR MODEL AND FILTERING APPROACH

The strapdown inertial error state variables and dynamics matrix are defined in Tables 1 and 2. These are coordinatized in the reference level wander azimuth coordinate system denoted by superscript L in Figure 1. In a strapdown system, the reference axes are computed with respect to body axes using quaternions. However, misalignments of these axes propagate just like platform axis misalignments, as indicated in Table 1. The accelerometer and gyro errors, denoted D and E, respectively, are defined in body axes as indicated by the matrices in Table 2. These do not couple into the level axes directly as they would in a platform system. Instead, the dynamics matrices of Table 2, are pre-multiplied by the direction cosine matrix from body to level axes to couple the errors into level axes, as shown in Table 1. The error model for a strapdown system is thus seen to be relatively simple, even though the mechanization is complex.

TABLE 1. ERROR STATES AND DYNAMICS MATRIX

$$\dot{\mathbf{x}}^T = \begin{bmatrix} \delta p_X & \delta p_Y & \delta p_Z & \delta v_X & \delta v_Y & \delta v_Z & \delta x & \delta y & \delta z & \delta^T & \mathbf{E}^T \end{bmatrix}$$

(NOTE - Blank elements den. as zero)

δp_X				1						
δp_Y	+	+			1					
δp_Z						1				
δv_X							$-A_Z$	A_Y		
δv_Y	+	+		+	+		A_Z	$-A_X$	$C_{B/D}^{L,B}$	
δv_Z							$-A_Y$	A_X		
δx										
δy										
δz										
δ										
\mathbf{E}										

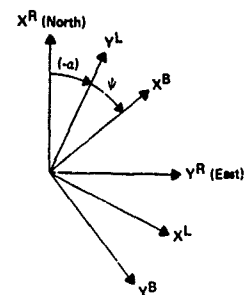


Figure 1. Coordinate Systems

TABLE 2. STRAPDOWN INSTRUMENT ERROR STATES AND DYNAMICS MATRICES IN BODY AXES

ACCELEROMETER ERROR STATES																		
BIAS ERRORS			SCALE FACTOR AND MISALIGNMENT ERRORS									NONLINEARITIES						
$\delta^T =$	δx_0	δy_0	δz_0	δx_{AX}	δx_{AY}	δx_{AZ}	δy_{AX}	δy_{AY}	δy_{AZ}	δz_{AX}	δz_{AY}	δz_{AZ}	δx_{AX2}	δy_{AY2}	δz_{AZ2}	δx_{AX3}	δy_{AY3}	δz_{AZ3}
$\delta^B =$	1	+	+	A_X^B	A_Y^B	A_Z^B	+	+	+	+	+	+	$(A_X^B)^2$	+	+	$(A_X^B)^3$	+	+
		1		+	+	+	A_X^B	A_Y^B	A_Z^B	+	+	+	+	$(A_Y^B)^2$	+	+	$(A_Y^B)^3$	+
			1							A_X^B	A_Y^B	A_Z^B		+	$(A_Z^B)^2$		+	$(A_Z^B)^3$
	I			DA									DA2			DA3		

GYRO ERROR STATES																		
BIAS ERRORS			SCALE FACTOR AND MISALIGNMENT ERRORS									MASS UNBALANCE						
$\delta^T =$	δx_0	δy_0	δz_0	δx_{WX}	δx_{WY}	δx_{WZ}	δy_{WX}	δy_{WY}	δy_{WZ}	δz_{WX}	δz_{WY}	δz_{WZ}	δx_{AX}	δx_{AY}	δx_{AZ}	δy_{AY}	δz_{AZ}	δz_{AZ}
$\delta^B =$	1	+	+	W_X^B	W_Y^B	W_Z^B	+	+	+	+	+	+	A_Y^B	A_X^B	+	+	+	+
		1		+	+	+	W_X^B	W_Y^B	W_Z^B	+	+	+	+	+	A_X^B	A_Y^B	+	+
			1							W_X^B	W_Y^B	W_Z^B	+	+	+	A_X^B	A_Y^B	A_Z^B
	I			EW									EA					

The usual approach to the initialization, alignment, and calibration problem is to use a single, large Kalman filter. This filter would model all the important states from Tables 1 and 2 together with two extra states for GPS user clock bias and drift. As shown in Reference 3, this could result in up to 40 states. Instead, the approach used here will be to partition the filter into many smaller filters. Especially important are fourth order filters used during straight and level flight. By observing changes in the estimated states of these filters before and after maneuvers, additional information is obtained for aligning the system and calibrating all significant instrument errors.

3. FILTERS FOR STRAIGHT AND LEVEL FLIGHT

The error model for straight and level flight is given in Figure 2 as a control system diagram. Although it describes the same model as the dynamics matrix of Table 1, this diagram is a more useful form for understanding the control loops of the Kalman filter. These loops are shown in Figure 3. Observe that the gyro and accelerometer error states from Figure 2 have been combined with other error states to obtain equivalent acceleration error and acceleration rate error states. The optimal gains for these loops are time varying and will be derived in Section 6.

It is observed that these filters have the same structure as stored gain fourth order ground alignment filters, where the position reference is zero. Such filters perform better than third order ground alignment filters which use a zero velocity reference for the observation. For the same reason, to be explained shortly, a position reference is used here rather than a velocity reference. However, as in ground alignment, this reference is not absolute position, as would be obtained from GPS code tracking loops. Instead, it is a relative position obtained by integrating delta range information from the carrier tracking loops.

To determine this relative position, at least four carrier tracking loops must be phase-locked with no cycle slippage. The use of rate aiding will reduce the probability of cycle slipping. By matrix inversion, the delta ranges are converted directly into delta position, as indicated in Table 3. In effect, these delta positions are then accumulated to obtain the relative position reference. To avoid scaling problems, the delta positions are differenced with deltas in estimated position before accumulation. The difference is then accumulated to obtain the position error observation, as indicated in Figure 4.

The reason this approach is more accurate than using delta range observations is shown in Table 4. The average velocity error of the estimated position is proportional to the phase error divided by the averaging time. The phase error is bounded by about 0.1 feet (60 degrees) since there is no cycle slippage. The averaging time is a function of the time constant of the Kalman filter, which is about 10 seconds. The average rate of change of position error is then less than 0.01 feet/second multiplied by a geometric

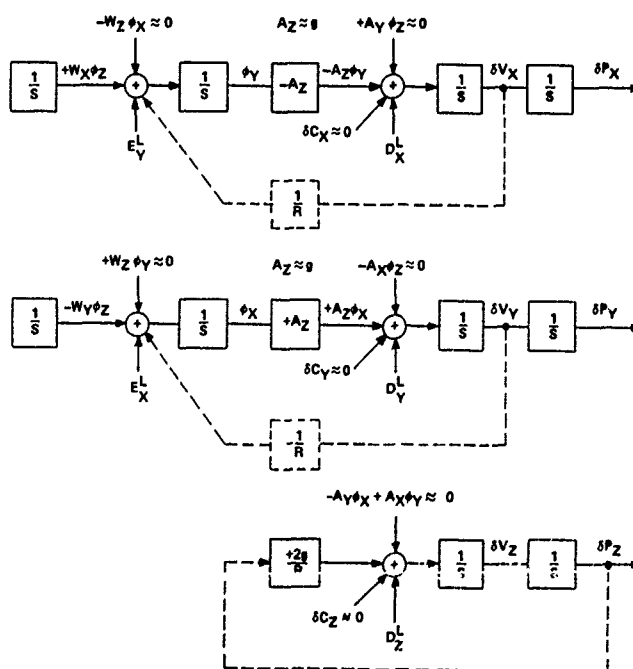


Figure 2. Approximate Error Models for Straight and Level Flight

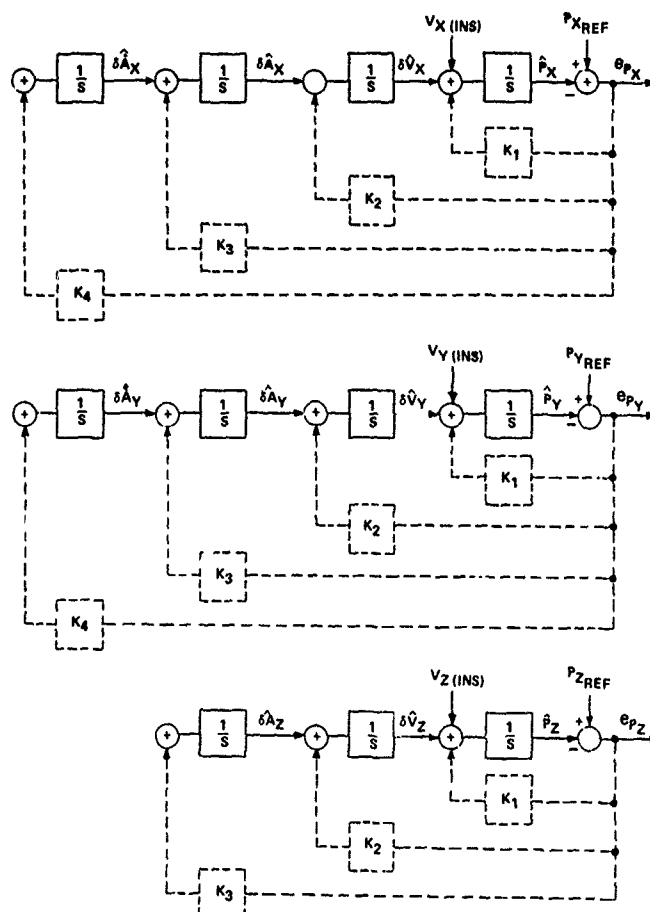
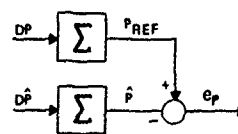


Figure 3. Kalman Filter Solutions to Error Models In Straight and Level Flight

TABLE 3. SOLUTION FOR DELTA POSITION

4x1	4x4	4x1
$\begin{bmatrix} DR_1 \\ DR_2 \\ DR_3 \\ DR_4 \end{bmatrix}$	$G = \begin{bmatrix} g_{11} & g_{12} & g_{13} & 1 \\ g_{21} & g_{22} & g_{23} & 1 \\ g_{31} & g_{32} & g_{33} & 1 \\ g_{41} & g_{42} & g_{43} & 1 \end{bmatrix}$	$\begin{bmatrix} DPX \\ DPY \\ DPZ \\ DBU \end{bmatrix}$
Observation Equation:		
$DR = G \cdot DP$		
Solution at 1 Hz.		
$DP(t) = [G(t)]^{-1} \cdot DR(t), t = 1, 2, \dots, T$		

a. Desired Mechanization:



b. Actual Equivalent Mechanization:

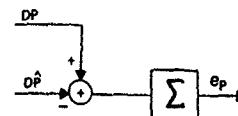


Figure 4. Mechanization for Computing Relative Position Error from Delta Position

TABLE 4. AVERAGE VELOCITY ERROR OVER T SECONDS

Since $[G]$ is approximately constant,

$$\begin{aligned} \delta \left(\sum_{t=1}^T DP(t) \right) &\approx [G(T)]^{-1} \cdot \delta \left(\sum_{t=1}^T DR(t) \right) \\ \delta \dot{P}_{AVE} &= \frac{\delta(\sum DP)}{T} \approx [G]^{-1} \frac{\delta(\sum DR)}{T} \\ &= [G]^{-1} \frac{e_{PHASE}}{T} \end{aligned}$$

dilution factor. By comparison, the use of delta range observations requires an interval of 1 second or less, which gives a velocity error of 0.1 feet/second multiplied by the same dilution factor. In this case the time constant is shorter, and the errors are reduced by the square-root of the averaging time, rather than by the averaging time itself, as in the approach proposed here.

4. ALIGNMENT AND CALIBRATION

Alignment and calibration are performed using outputs from the filters of Figure 3. These outputs are the estimated velocity errors, acceleration errors, and acceleration rate errors in the X, Y channels. From the error model of Figure 2, the outputs are related to the error states as indicated in Table 5. As shown, the filter outputs are each combinations of instrument error states and reference axis misalignments. Since the missile will maneuver or accelerate during the midcourse phase, these combinations will not remain constant. It is therefore necessary to separately estimate the misalignments and instrument errors during the alignment and calibration phase before GPS is jammed. This separation requires maneuvers similar to those to be encountered during the midcourse phase. To see how these combinations change during a turn, consider the body coordinate systems before and after the turn, as defined in Figure 5. Since the filters shown in Figure 3 are open loop filters, the estimated errors before and after the turn can be observed independently. The changes in these estimates are denoted by a capital delta.

TABLE 5. FILTER OUTPUTS IN STRAIGHT AND LEVEL FLIGHT

$$\text{Define } C_B^L = \begin{bmatrix} \sin \psi & \cos \psi \\ \cos \psi & -\sin \psi \end{bmatrix}$$

$$\begin{bmatrix} \delta \hat{A}_X \\ \delta \hat{A}_Y \end{bmatrix} = C_B^L \cdot \begin{bmatrix} D_X \\ D_Y \end{bmatrix} + \begin{bmatrix} -g\phi_Y \\ +g\phi_X \end{bmatrix}$$

$$\begin{bmatrix} +\delta \hat{A}_Y/g \\ -\delta \hat{A}_X/g \end{bmatrix} = \begin{bmatrix} \delta \omega_X \\ \delta \omega_Y \end{bmatrix} = C_B^L \cdot \begin{bmatrix} E_X \\ E_Y \end{bmatrix} + \begin{bmatrix} -W_Y\phi_Z \\ +W_X\phi_Z \end{bmatrix}$$

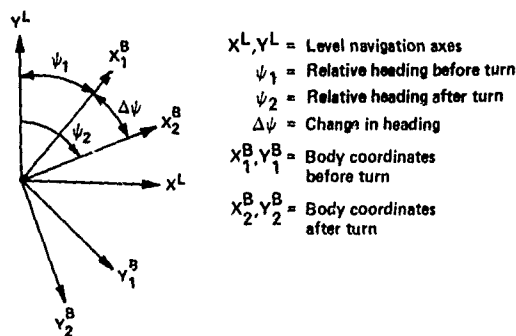


Figure 5. Coordinate Systems for Accelerometer and Gyro Bias Estimation

With these definitions, the alignment and calibration solutions are as given in Table 6. The initial leveling solution is obtained by solving the filter output equations of Table 5 with accelerometer errors D set to zero. After the first maneuver, the azimuth misalignment error is determined by observing changes in velocity error perpendicular to

TABLE 6. SOLUTION FOR STATE VARIABLES

Conventional Leveling Solution:

$$\begin{bmatrix} -g\phi_Y \\ +g\phi_X \end{bmatrix} = \begin{bmatrix} \delta \hat{A}_X \\ \delta \hat{A}_Y \end{bmatrix}$$

Conventional Maneuver Alignment Solution:

$$\hat{\phi}_Z = \frac{\Delta V_Y \cdot \Delta \hat{V}_X - \Delta V_X \cdot \Delta \hat{V}_Y}{(\Delta V_X^2 + \Delta V_Y^2)}$$

Accelerometer and Gyro Bias Estimates:

$$\text{Define } \Delta C_B^L = \begin{bmatrix} (\sin \psi_2 - \sin \psi_1) & (\cos \psi_2 - \cos \psi_1) \\ (\cos \psi_2 - \cos \psi_1) & -(\sin \psi_2 - \sin \psi_1) \end{bmatrix}$$

$$\begin{bmatrix} \hat{D}_X \\ \hat{D}_Y \end{bmatrix} = [\Delta C_B^L]^{-1} \cdot \begin{bmatrix} \Delta \delta \hat{A}_X \\ \Delta \delta \hat{A}_Y \end{bmatrix}$$

$$\begin{bmatrix} \hat{E}_X \\ \hat{E}_Y \end{bmatrix} = [\Delta C_B^L]^{-1} \left\{ \begin{bmatrix} \Delta \delta \hat{A}_Y/g \\ -\Delta \delta \hat{A}_X/g \end{bmatrix} - \begin{bmatrix} -\Delta W_Y \hat{\phi}_Z \\ +\Delta W_X \hat{\phi}_Z \end{bmatrix} \right\}$$

the velocity change. This is obtained as a vector cross product, as shown. To estimate accelerometer and gyro bias errors, the change in the 2x2 direction cosine matrix of Table 5 is observed and its inverse is computed. The X and Y accelerometer and gyro bias errors can then be determined as shown in Table 6. These solutions assume no change in the misalignments due to the maneuver. The estimate for level axis misalignments can now be corrected, since it was based on assuming accelerometer errors D were zero. Finally, the Z accelerometer bias error is estimated directly from the vertical channel filter of Figure 3, while the Z gyro bias error is determined by comparing azimuth misalignment corrections from maneuvers separated by a time interval.

The foregoing discussion has explained how the simple filters of Figure 3 can be used to analytically determine all the misalignments and instrument bias errors normally included in a 17-state GPS/INS filter. However, it was shown in Reference 3 that when maneuvers are included in the midcourse phase, many additional instrument errors must be calibrated in order to achieve the desired accuracy goal. These include gyro scale factor and misalignment errors, gyro mass unbalance (g-sensitive) errors, accelerometer misalignment errors and accelerometer Z-axis scale factor errors. The excessive position error contribution of these sources, even when using a 17-state filter was shown in Reference 3 by using a covariance analysis based on a 6 degree-of-freedom trajectory simulation.

For a winged missile, the largest accelerations are along the Z body axis which determines the g-sensitive errors requiring calibration. The accelerometer Z-axis scale factor errors can be calibrated separately by observing change in velocity error in the direction of the velocity change, rather than in the perpendicular direction as used to calibrate azimuth misalignment error. The only accelerometer misalignments requiring calibration are the X and Y accelerometer misalignments toward Z, denoted DXAZ and DYAZ. These can be avoided by simply re-defining the body axes so that the X and Y axes lie in the plane of the X and Y accelerometer sensitive axes. No other accelerometer errors in Table 2 require calibration, except the bias errors.

This leaves the gyro scale factor, misalignment, and g-sensitive errors. Since large accelerations lie along the body Z-axis, the gyros can be arranged so that the only g-sensitive drift requiring calibration is the Z-gyro drift due to Z-acceleration. In addition, all gyro scale factor and misalignment errors require calibration. This makes a total of 10 gyro error sources to be calibrated in addition to the two bias errors previously considered. To calibrate these 10 additional error states, the solutions to the filter output equations in Table 6 must be modified. The previous solution for accelerometer bias errors was obtained by assuming no change in X, Y misalignments due to the maneuver. The 10 additional gyro error sources will contribute changes in the X and Y misalignment angles, producing the observation equations of Table 7. The addition of the 10 gyro error states to the 2 accelerometer biases results in 12 states. These observation equations are used after each maneuver to estimate the 12 error states in a separate 12-state filter. The errors are treated as fixed bias errors with no dynamics matrix. This solution can be performed at a slow rate in a background loop since the observations occur only after each maneuver, and since there is no dynamics matrix.

TABLE 7. OBSERVATION EQUATIONS FOR CALIBRATING GYRO PARAMETERS

$$\begin{bmatrix} \Delta \hat{A}_X \\ \Delta \hat{A}_Y \end{bmatrix} = \begin{bmatrix} 2 \times 1 & 2 \times 2 \\ \Delta C_B^L & D_X^B \\ & D_Y^B \end{bmatrix} + \begin{bmatrix} 0 & -g & 0 \\ g & 0 & 0 \end{bmatrix} \cdot \left(\int_{t_1}^{t_2} C_B^L \cdot \begin{bmatrix} 3 \times 15 \\ [EW, EA] \end{bmatrix} dt \right) \begin{bmatrix} E_{IW} \\ E_{IA} \end{bmatrix}$$

5. CARRIER LOOP RATE AIDING

Reference 3 showed that position accuracy can be maintained for at least 5 minutes after jamming, provided the inertial system is accurately initialized, aligned, and calibrated before jamming occurs. This is illustrated in Figure 6, where the principal error sources are each selected to yield less than 10 feet position error per channel 5 minutes after jamming. However, it is seen that position errors grow rapidly 10 minutes after jamming because of gyro noise. It is therefore desirable to delay jamming as long as possible in order to shorten the time during which inertial errors can propagate. This can be accomplished by using the inertial system to rate-aid the tracking loops so that their bandwidths can be lowered without increasing errors due to dynamics. The code loop can avoid loss of lock longer than the carrier loop and is easier to rate-aid. However, velocity errors will begin to increase at the time the carrier loop loses lock. Therefore, carrier loop rate-aiding will be considered here. In a high dynamic environment, a carrier loop bandwidth of about 30 Hz is required without rate-aiding. Assuming a good user clock, the bandwidth can be lowered to 1 Hz with proper rate-aiding, thus improving the anti-jam performance by roughly 15 dB.

Despite this potential performance improvement, early proposals (Refs. 5,6) to rate-aid the carrier loop were not implemented. There are two reasons for this. The first is that the time lags that can be tolerated in the inertial data are much smaller than originally anticipated. The second is that the effects of relative motion between the INS and GPS antenna are more than anticipated. To see this, the phase error in the

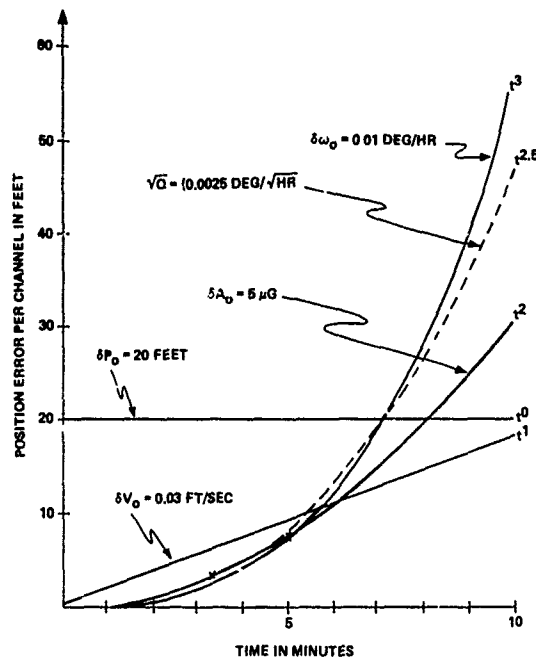


Figure 6. Error Growth Due To Initial Errors

tracking loop due to dynamics alone should not exceed 30 degrees. Since the wavelength is only 0.623 feet, this is only 0.05 feet error. Assuming a sinusoidal jerk of 5 g/sec at 3 radians/sec, a delay of only one millisecond in the aiding information causes a 20 degree phase error in the tracking loop. Also, with a sinusoidal roll of 60 degrees peak at 10 radians/sec and a 2 foot lever arm a delay of one millisecond in lever arm compensation causes a 20 degree phase error. It will be shown that delays in the aiding information can easily be as large as 40 milliseconds. With no compensation for these delays, or for the relative antenna motion, attempts at carrier rate aiding will not succeed.

Figure 7 illustrates the delays to be expected in the aiding information. Beginning at the right, the numerically controlled oscillator must advance the phase in the tracking loop over the measurement interval. The amount of the advance must be computed ahead of time, at time t as shown. Because of delays in computation and transmission, the most recent inertial measurement unit (IMU) sample is not yet available for use at time t . The next most recent sample is delayed an additional amount equal to the sampling interval of the inertial aiding information. The total time delay, referred to as extrapolation time, is the interval from when this aiding information was sampled to the

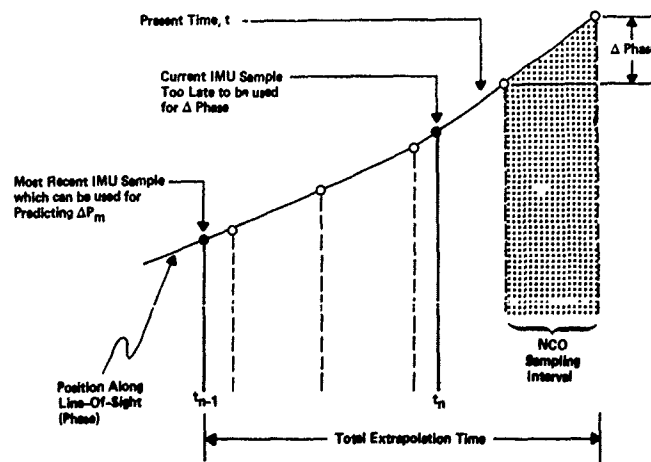


Figure 7. Extrapolation Time Requirements

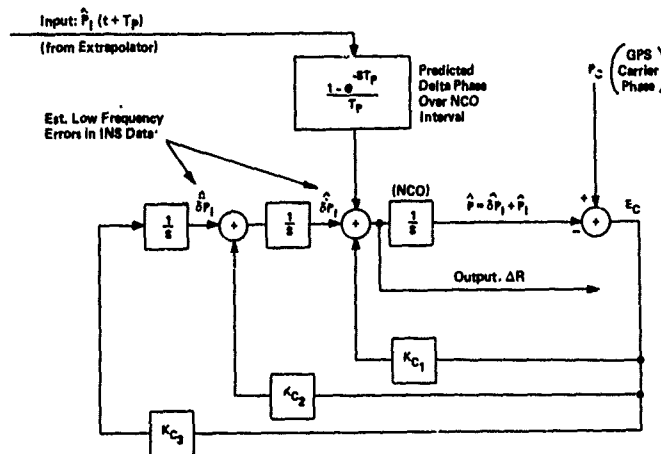


Figure 10. Low Bandwidth GPS Carrier Tracking Loop

tracking computer may have little time for extra computations, each extrapolator requires very little duty cycle since only the summations of acceleration into velocity and velocity into position are performed at the high rate.

Because of the time lags, it is necessary for the extrapolator outputs to apply at a time in advance of real time. Therefore, before the sampled extrapolator position is compared with the inertial sample, it is deliberately delayed by the sum of its prediction time and the lag in the inertial data as shown in Figure 9. After an additional delay, the updates to the extrapolator states are computed. By this time they are behind the extrapolator time by the total time T which is the sum of the prediction time, the inertial data lag, and the update computation lag. The optimal updates to the extrapolator states are then predicted from the delayed updates by multiplying by the transition matrix of the extrapolator, evaluated at time T as shown. It is assumed that all delays are known accurately, so that T is known to within a millisecond. This requires using a timing signal between the inertial system and GPS tracking computer.

The optimal steady-state gains for the extrapolator are determined by simulation and are stored as constants. These are based on a model of the true inertial position as being caused by white noise at the acceleration rate level (jerk) followed by three integrators in series. In Reference 3, these gains were determined and the extrapolator performance was tested with an input from a simulated missile autopilot response. The autopilot input was a 5g step acceleration command which resulted in a maximum jerk of almost 14 g's/sec. The error in the extrapolator was found to increase with total extrapolation time as shown in Figure 11. Even with this severe dynamics and a total extrapolation time of 40 milliseconds the extrapolation error is only 0.05 feet, or about 30 degree phase error. This shows that a simple extrapolator can be used to overcome the effects of large time lags in carrier loop aiding data.

6. OPTIMAL GAINS

This section will address a general class of Kalman filtering problems illustrated in Figure 12. The solution is useful in solving many practical filtering problems that arise in practice. In particular, the solutions apply to simple alignment or navigation

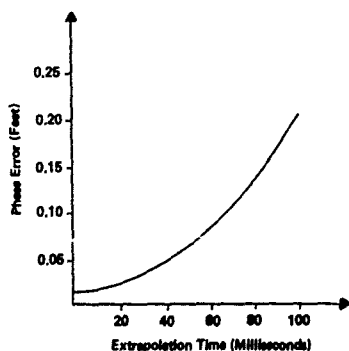


Figure 11. Extrapolation Error

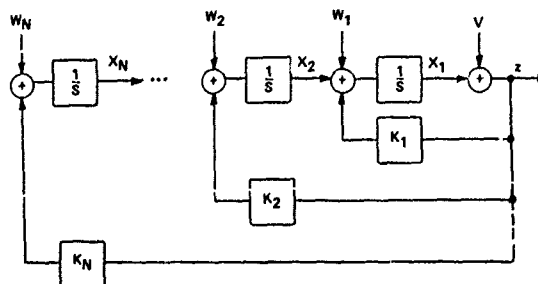


Figure 12. Kalman Filtering of States in Series

filters of the type shown in Figure 13. The solutions also apply to optimal tracking loops as shown in Figure 14. These include GPS carrier tracking loops of first, second, and third order. One solution to such problems is to determine the gains K in real time by the recursive Kalman filter algorithm. This means that neither the gains nor the resulting performance are known generally in advance of the real application. Instead, a few isolated cases are simulated and flight tested in the hope that satisfactory performance in these cases is representative of the range of cases to be encountered. This approach differs from that used in classical control, when both the gains and performance over an infinite set of circumstances are known in advance of the actual application.

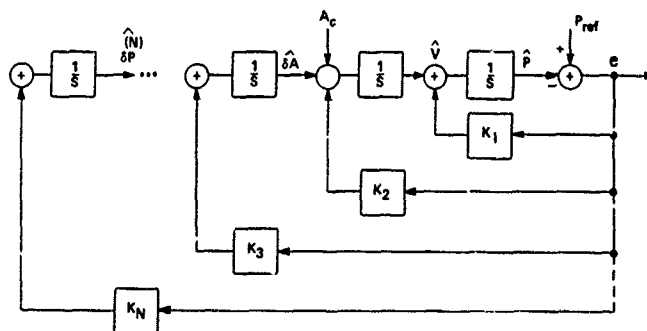


Figure 13. An N^{th} Order Navigation Filter

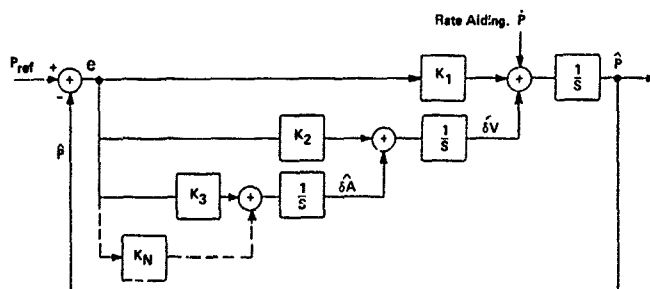


Figure 14. General N^{th} Order Tracking Loop

The problem is compounded by the fact that the optimal gains vary erratically and are not simple functions of time. Reference 3 shows that even a simple third order problem leads to gains which change by orders of magnitude and include many inflection points. These inflection points are not easy to distinguish from undesired oscillations caused by numerical problems in the algorithm. At the same time, the real oscillations from the inflection points are sensitive to the assumed initial conditions. These problems are especially troublesome when an attempt is made to avoid a real-time Kalman filter by using pre-stored gains based on simulations, as for the partitioned filters considered here.

Because of these difficulties it is desirable to determine the optimal gains analytically as far as possible. The simplest case represented by Figure 12 is a single state variable with observation noise but no process noise. The solution for this case gives insight into the more general case. When the initial variance is infinite, the solution is given in Table 8 and is mechanized as an open loop continuous Kalman filter in Figure 15. The optimal gain is $1/t$. Even this simple case would be awkward to store as a gain schedule if the analytic solution were not known. To simplify the gain scheduling problem, the gain is determined by multiplying it by a constant factor at time intervals which increase at a constant ratio. In this way, whenever t doubles, K is reduced by half. To understand the procedure, the log of K is a linear function of the log of t . Therefore, K vs. t on log-log scaled plots appears as a straight line with slope -1 , as shown in Figure 16.

The gain decreases as $1/t$ only until time τ , when it levels off at a constant value $1/\tau$. This illustrates the next simplest case from Figure 12 where there is process noise w in addition to observation noise v . As a result, the gain decreases as $1/t$ initially to rapidly reduce the estimation error caused by large initial condition errors. But when these errors are reduced sufficiently, the plant noise becomes significant, and the gain is not reduced further. The solution is then divided into a transient segment where the gain is $1/t$, and a steady-state segment where the gain is $1/\tau$. Since both segments appear as straight lines on the log-log scale, the gain is easily computed as described previously. That is, on each segment the gain is multiplied by a constant factor at time intervals which increase at a constant ratio.

TABLE 8. RECURSIVE ESTIMATE OF ERROR STATE WITH NOISE IN OBSERVATION

Observation.
$Z_i = X + V_i$
Optimal Estimate of X After N Observations
$\hat{X}_N = \frac{N}{\sum_{i=1}^N} \frac{Z_i}{N}$
$= \frac{1}{N} Z_N + \frac{N-1}{N} \left(\frac{1}{N-1} \sum_{i=1}^{N-1} Z_i \right)$
$= \frac{1}{N} Z_N + \frac{N-1}{N} (\hat{X}_{N-1})$
$= \hat{X}_{N-1} + \frac{Z_N - \hat{X}_{N-1}}{N}$
$\frac{\hat{X}_N - \hat{X}_{N-1}}{dt} = \frac{(Z_N - \hat{X}_{N-1})}{Ndt}$
In Limit.
$\dot{\hat{X}} = K \cdot (Z - \hat{X})$
When
$K = \frac{1}{Ndt} \rightarrow \frac{1}{t} \text{ as } dt \rightarrow 0, N \rightarrow \infty$

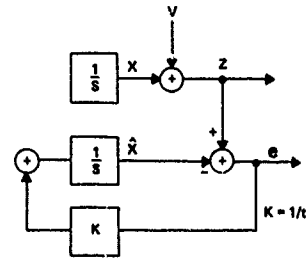


Figure 15. Mechanization Diagram for Recursive Estimate of Error State

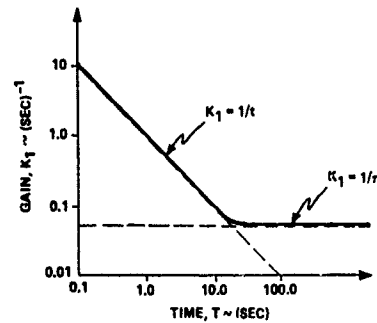


Figure 16. Optimal Gain for Estimating Single Error State

To determine the time constant, the steady-state solution to the problem of Figure 12 will be determined by solving the matrix-Riccati equation in Table 9. The dynamics matrix F, observation matrix H, process noise Q, and observation noise R for the general case of Figure 12 are given in Table 10 and are substituted into Table 9. It is assumed now that only process noise for $i=N$ is non-zero. The analytic solutions in this case are given in Table 11 for $N=1,2,3$. The pole-zero locations of the filters are shown in Figure 17. For $N=2$, the solution required solving three equations in three unknowns, which are the upper triangular elements of the covariance matrix. For $N=3$, this required solving six equations in six unknown covariance elements. Since the equations are non-linear, this becomes increasingly difficult. However, the solution for all N can be obtained by observing the pattern for the normalized gain ratios as defined in Table 12.

TABLE 9. STEADY-STATE KALMAN FILTER EQUATIONS

Steady-State Matrix Riccati Equation:
$FP + (FP)^T + Q = PH^T R^{-1} HP$
Steady-State Gains
$K = PH^T R^{-1}$

TABLE 10. MATRICES FOR STATES IN SERIES

Dynamic Matrix:	Identity Matrix:
$N \times N$	$(N-1) \times (N-1)$
$F = \begin{bmatrix} 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & 0 \end{bmatrix}$	$I_{N-1} = \begin{bmatrix} 1 & 0 & 0 & \dots & 0 \\ 0 & 1 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & 1 \end{bmatrix}$
Observation Matrix:	Process Noise Matrix:
$1 \times N$	$Q = \begin{bmatrix} Q_1 & 0 & 0 & \dots & 0 \\ 0 & 0 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & Q_N \end{bmatrix}, Q_i = \sigma_{w_i}^2 T$
$H = [1, 0, \dots, 0]$	1×1
Observation Noise:	$R = (\sigma_v^2 \cdot T)$

TABLE 11. STEADY-STATE KALMAN FILTERS FOR STATES IN SERIES AND Nth ORDER PROCESS NOISE

$Q_N = Q \quad (Q_i = 0, i < N)$			
Order	Time Constant	Gains	Covariance
N=1	$\tau = \left(\frac{R}{Q}\right)^{1/2}$	$K_1 = \frac{1}{\tau}$	$P_{11} = Q\tau$
N=2	$\tau = \left(\frac{R}{Q}\right)^{1/4}$	$K_1 = \frac{\sqrt{2}}{\tau}$ $K_2 = \frac{1}{\tau^2}$	$P_{11} = \sqrt{2}Q\tau^3, P_{12} = Q\tau^2$ $P_{22} = \sqrt{2}Q\tau$
N=3	$\tau = \left(\frac{R}{Q}\right)^{1/6}$	$K_1 = \frac{2}{\tau}$ $K_2 = \frac{2}{\tau^2}$ $K_3 = \frac{1}{\tau^3}$	$P_{11} = 2Q\tau^5, P_{12} = 2Q\tau^4, P_{13} = Q\tau^3$ $P_{22} = 3Q\tau^3, P_{23} = 2Q\tau^2$ $P_{33} = 2Q\tau$

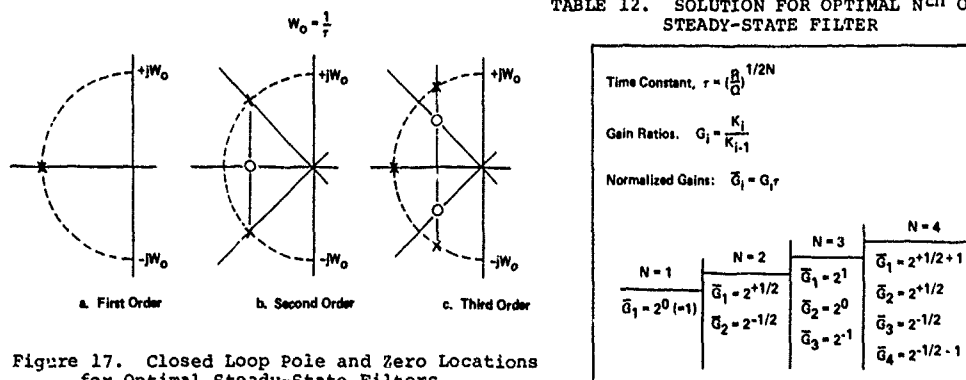
TABLE 12. SOLUTION FOR OPTIMAL Nth ORDER STEADY-STATE FILTER

Figure 17. Closed Loop Pole and Zero Locations for Optimal Steady-State Filters

These normalized gain ratios are also useful for studying the transient response, since their characteristics with log-log scaling are easily approximated by piecewise linear segments. The transient and steady-state solutions vs. normalized time are shown in Figure 18 for various cases where initial conditions are infinite. The second and fourth cases shown apply where only process noise enters the highest order node and the entire filter reaches steady state. Observe that the steady-state normalized gains are equally spaced, since successive gains are always in the ratio 2:1 as seen in Table 12. The first, third, and fifth cases have an additional higher order state entering this node.

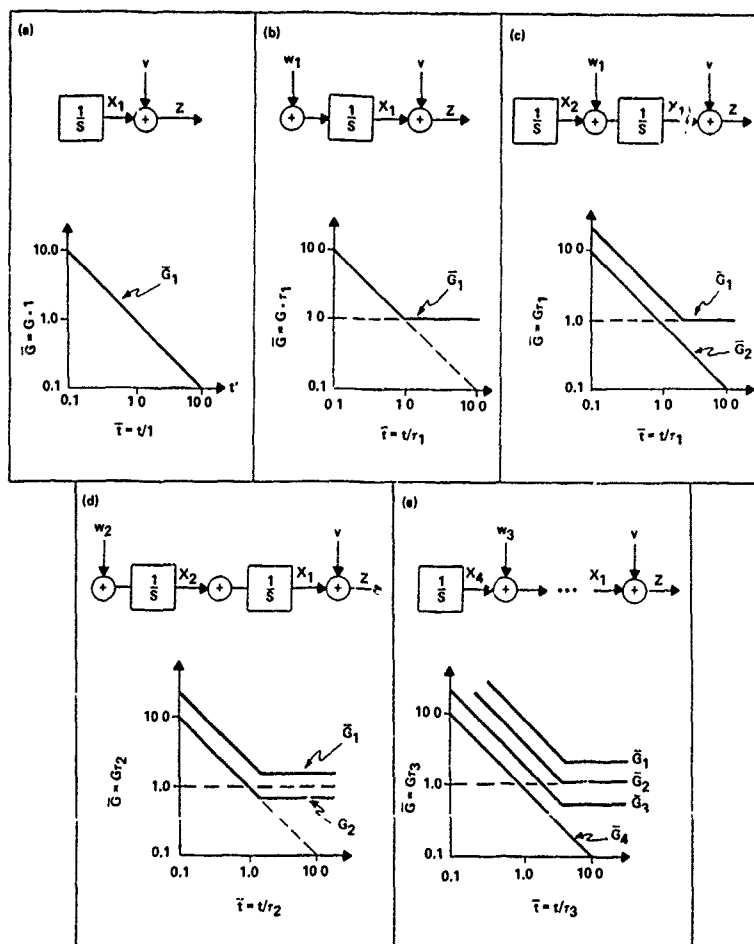


Figure 18. Normalized Gain Schedules

The gain for this state continues decreasing as $1/t$, while the gains for all lower order states reach steady-state. This steady-state filter produces a single continuous output for observing the highest order state. This is similar to the first case studied for $N=1$, except that the observation noise is replaced by the PSD of the total output noise of the steady-state filter. The fifth case gives the solution to the level axis filters of Figure 3, where the process noise is the gyro white noise and the observation noise is the carrier loop error.

If there is more than one source of process noise, the predominant source is first assumed to be the only non-zero source, and the steady-state solution is determined as just described. The steady-state effects of process noise at lower order nodes is then determined by contour integration, using the original steady-state filter to determine the transfer function from input noise source to output state estimate. The contour integral of the power spectral density for the output state variance is given by the sum of the residues in the half-plane. It is evaluated in Table 13 for arbitrary filter transfer functions up to order 3. The evaluation of this integral may show that the variance in the estimated states is larger than that given by Table 11. The predominant source of process noise must then be assumed to be the lower order source, and the steady-state filter for this source must be determined. The effect of less than predominant process noise at higher order nodes must then be considered. For example, assume there is a weak process noise at a fourth order node not shown in Figure 18, case (e). The fourth gain would not decrease indefinitely, but would eventually level off at a low value determined by a very long time constant. This time constant is given by the first order case in Table 11, except that the observation noise is the total output PSD of the third order steady-state filter assuming only the original process noise at the third node.

The foregoing filter solutions were based on the assumption that the initial conditions were infinite. In most cases, the corresponding filter can be used with finite initial conditions even though the solution is not optimal for that case. The variances during the initial transient period will be slightly larger than optimal, but the steady-state values will still be optimal. When finite initial conditions must be considered, the effect is as shown in Figure 19. This shows the effect of a finite initial variance for the fourth state. The corresponding gain does not start off at the large initial values for $1/t$. Instead, it increases from zero until it overshoots the $1/t$ schedule and then gradually decays toward this curve. The effects on gains of lower order is to produce oscillations as shown. Figure 18 should then be compared with Figure 18(e) which was based on infinite initial conditions. The effect of finite initial conditions of lower order states is less severe, as illustrated for the first gain in Figure 19.

TABLE 13. EVALUATION OF STEADY-STATE ESTIMATION ERRORS DUE TO NOISE

<p>Process Noise, w_i (PSD = Q_i)</p> <p>Filter</p> <p>Estimated State, X_i</p> <p>$G(S)$</p>	
<p>Filter Transfer Function:</p> $G(S) = \frac{c_{n-1}S^{n-1} + \dots + c_0}{d_nS^n + d_{n-1}S^{n-1} + \dots + d_0}$	
<p>Steady-State Error Variance:</p> $\sigma_{X_i}^2 = I_n \cdot Q_i, \text{ where } I_n = \frac{1}{2\pi j} \int_{-\infty}^{+\infty} G(S) G(-S) dS$	
<p>Evaluation of Integral:</p>	
$n=1$	$I_1 = \frac{c_0^2}{2d_0d_1}$
$n=2$	$I_2 = \frac{c_1^2d_0 + c_0^2d_2}{2d_0d_1d_2}$
$n=3$	$I_3 = \frac{c_2^2d_0d_1 + (c_1^2 - 2c_0c_2)d_0d_2 + c_0^2d_2d_3}{2d_0d_1(-d_0d_2 + d_1d_2)}$

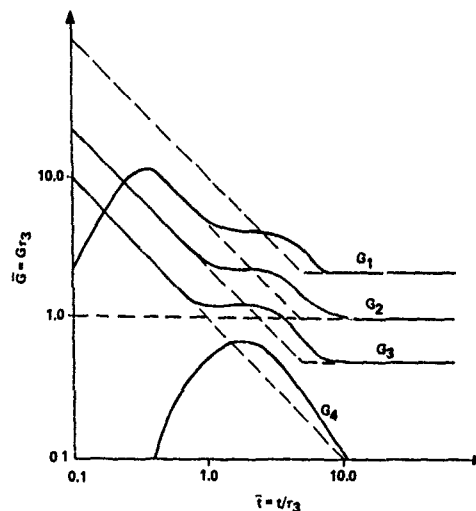


Figure 19. Effects of Finite Initial Conditions

7. CONCLUSIONS

It has been shown that GPS/INS integration for optimal performance after jamming can be achieved by use of partitioned low order filters. In particular, fourth order filters are used for measuring changes in velocity, acceleration, and acceleration rate errors before and after maneuvers. These measurements are used to obtain analytic solutions to the in-flight alignment and instrument biasing problem normally solved by a 17-state Kalman filter. Also, third order filters are used for extrapolating inertial information to aid the carrier tracking loop.

The optimal gains for a general class of filtering problems was also determined. These solutions apply to the partitioned filters used here. They make it possible to compute the gains in real time by simple multiplications, rather than by solving a complicated recursive Kalman filter algorithm. Since the gains and other solutions are determined analytically, performance can be analyzed and tested in advance of the real-time application. This may be more reliable than depending on a real-time recursive Kalman solution not known in advance.

REFERENCES

1. Kao, M.H., and Eller, D.H., "Multiconfiguration Kalman Filter Design for High-Performance GPS Navigation", IEEE Trans. Automatic Control, Vol. AC28, No. 3, 1983, pp. 304-314.
2. Sturza, M.A., Brown, A.K., and Kemp, J.C., "GPS/AHRS: A Synergistic Mix", IEEE NAECON, 1984, pp. 339-348.
3. Diesel, J.W., "Integration of GPS/INS for Maximum Velocity Accuracy", Proceedings of the National Technical Meeting of the ION, Anaheim, CA, January 20, 1987, pp. 56-65.
4. Buechler, D., and Foss, M., "Integration of GPS and Strapdown Inertial Subsystems into a Single Unit", Navigation, Journal of the ION, Vol. 34, No. 2, Summer, 1987, pp. 140-159.
5. Hemesath, N.B., "Performance Enhancements of GPS User Equipment", Global Positioning System, Institute of Navigation, Washington D.C., 1980, pp. 107-108.
6. Cox, D.B., "Integration of GPS with Inertial Navigation Systems", Global Positioning System, Institute of Navigation, Washington, D.C., 1980, pp. 148-150.

2-D AND 3-D CHARACTERIZATIONS OF GPS NAVIGATION SERVICE

P. Massatt, W. Rhodus, and K. Rudnick

The Aerospace Corporation

The quality of world-wide GPS navigation service is frequently described by plotting on a world map background 2-dimensional regions (latitude, longitude) on the earth's surface where navigation is degraded (typically, $PDOP > 6$ or $HDOP > 4$) for some period of time - over the span of a day. Such depictions provide little information concerning when such navigation degradations occur or their duration. Computationally efficient algorithms are presented for computing the locations in space and time of these navigation degradations. We also present quick geometric formulas for calculating the dilutions of precision (GDOP, PDOP, HDOP and TDOP). A data base is generated with this approach on a mainframe computer and then post-processed on a high resolution color graphics workstation to produce 3-dimensional plots (latitude, longitude, and time) which insightfully characterize GPS navigation service and its evolution in time. Degradation regions are readily isolated with their time of occurrence, duration, and location clearly defined. Also given are 2-dimensional plots which display location and time of poor coverage permitting an analysis of the quality of coverage from another perspective. Results are provided for the current 7 operational Block I satellites, a 9 satellite constellation projected for 1989 when 4 Block II SVs will have been added, the GPS baseline 21 SV constellation, and a revised 21 SV constellation under consideration.

SECTION 1. INTRODUCTION

The Global Positioning System (GPS) is a navigation and time transfer satellite system designed to provide continuous 3-dimensional position, velocity, and U.S. Naval Observatory Coordinated Universal Time (UTC) to users world-wide. Currently, there are 7 operational space vehicles (SVs) on orbit and revised launch schedules now plan for 21 SVs to be deployed by 1992.

The quality of GPS navigation service has frequently been described by plotting on a world map background 2-dimensional regions (latitude, longitude) on the earth's surface where navigation is degraded due to poor relative geometry of the satellites for some period of time typically over the span of a day (see Figure 3). Such depictions do provide some information about the duration of navigation outages but very little data concerning their time of occurrence - critically needed information to many users. The main thrust of this paper is to describe how regions of degraded coverage can be completely characterized, and to present these data in a readily interpretable form.

Section 2 and the Appendix describe computationally efficient algorithms for calculating GDOP, PDOP, HDOP, and TDOP globally in time and space coordinates. These algorithms are used to construct an "outage" data base, depicting the occurrence in time and space of degraded GPS navigational capability. These data are routinely calculated accurate to within 1.0 degree in latitude/longitude and 10 seconds in time.

Two graphical approaches are used to display the inherent 3-D (latitude, longitude, time) extent of a GPS navigation outage. The first employs a pair of plots (latitude, time) and (longitude, time), and the second uses a high resolution color graphics workstation to post process the outage data base file to plot outages directly in 3-D. Both approaches insightfully characterize GPS navigation service and its evolution in time. Outage regions are readily isolated with their time of occurrence, duration, and location clearly defined.

The above techniques are then applied and results are provided in Section 3 for the current 7 operational Block I satellites, a 9 satellite constellation projected for 1989 when 4 Block II SVs will have been added, the GPS baseline 21 SV constellation, and a revised 21 SV constellation under consideration.

SECTION 2. APPROACH

The problem is to determine where on the surface of the earth and when in time a user can expect good navigation accuracy from the GPS service. We often use the geometric dilution of precision, GDOP, as a quantitative figure of merit for "good navigation". GDOP is the ratio of the r.m.s. position and clock error to the 1- σ satellite ranging error. GDOP is not the only figure of merit we may wish to use. Since the majority of users are more interested in the accuracy of their position than in the accuracy of determining their clock bias, the positional dilution of precision, PDOP (ratio of the r.m.s. position error to the 1- σ satellite ranging error), is the important figure of merit. Other users may find HDOP (horizontal dilution of precision), VDOP (vertical dilution of precision) or TDOP (temporal dilution of precision) as their required figure of merit. Refer to the Appendix for computationally efficient, geometric formulas for calculating the various dilutions of precision. Once the appropriate figure of merit is defined, then a more quantitative statement of the problem is:

Determine when and where on the surface of the earth a user can access GPS with a dilution of precision less than a given threshold.

Typically, we look for solutions to the problem under the constraints $PDOP < 6$. This translates to 1- σ user position errors on the order of less than 36 meters after factoring in the system requirement that the r.m.s. 1- σ satellite ranging error shall not exceed 6 meters. Currently, satellite ranging errors are averaging less than 4 meters.

The important thing to observe is that the solution to the problem is the depiction of a 3-dimensional region in the variables latitude, longitude, and time. We have chosen to approximate this 3 dimensional region to a rather fine granularity (typically 1 square degree on the surface of the earth and less than 10 seconds in time), store the region as a database on a computer,

then post-process the database using a variety of graphical operators to fully visualize the solution to the problem.

Building the Database

Let us consider the specific problem of determining when and where on the surface of the earth we have $PDOP > 6$ for a given fixed constellation of satellites. The method we outline here is based upon representing the earth as a discrete set of points. The points are placed on latitude bands (usually with one degree separation) with the number of longitude points on each band decreasing with the cosine of the latitude to place an "equal area" grid over the surface of the earth. For each selected point the time intervals when $PDOP > 6$ are computed within an input tolerance (typically 10 seconds) using a root-solving method. When done, we have a file consisting of one record per (latitude, longitude) point which looks like:

```
header (number of lat,lon points, initial
      .      ephemerides, etc.)
      .
      .
/lat,lon, time intervals where PDOP > 6
      .
      .
end-of-file
```

We call such a file a dopcube.

A few words should be mentioned here about the actual algorithm used for building each dopcube. Many things can be done to improve efficiency. Points to note are:

1. Efficient geometric calculations of GDOP, PDOP, etc. are used (see Appendix). These equations are mathematically equivalent to the formal definitions of the "DOPS". The method used to select the best 4 or 3 satellites to base a DOP calculation on at a fixed lat, lon grid point is just to run through satellite combinations until the DOP threshold is satisfied. Not all j combinations of n satellites need to be computed. Note that no approximations are used in this process.
2. For each fixed ground point and each combination of 4 satellites a root-solving method is used to find time intervals of coverage.
3. The option to look at coarse grid spacings first, then finer grid spacings later, is incorporated for efficiency.
4. The time intervals are computed on successive buildups of the constellation, adding one satellite at a time rather than looking at all combinations at once.

Making use of all such efficiencies, a typical 21 satellite computer run (2 pixel size, 10 sec time accuracy) to generate a dopcube will take about five minutes of CPU time on an IBM 3090 while a dopcube generation run for the

current NAVSTARs 3,4,6,8,9,10,11 may take as long as forty minutes. The poorer the quality of the navigation service, the longer it takes. This is somewhat surprising given that the quality of the service is inverse to the number of satellites.

Databases associated with PDOP and HDOP (requiring 3 satellites and an estimate of the altitude uncertainty) thresholds are the most commonly used in GPS applications. These measures of the quality of the relative geometry of satellites involved in a users navigation solution can be converted to an assessment of 2-D and 3-D navigation accuracies by the following relation:

$$ND \pm DOP * URE \quad (1)$$

where

URE is a constant 1- σ user ranging error for all satellites,

DOP is either the PDOP or HDOP threshold,

and

ND is the associated 2-D or 3-D accuracy depending on the choice for DOP,

For example, Figure 8, can be interpreted as depicting where navigation to within 24 meters in 3-D is possible under the assumption that the URE for all SVs is not greater than 4 meters.

Equation 1) follows from the factorization of the covariance matrix associated with a least squares solution with constant measurement weights. In practice, satellite UREs are variable. An extension to accommodate variable UREs has been implemented in Aerospace software. The associated generalizations of the formulae in the Appendix will appear in a forthcoming paper.

3-D Dopcube Display

Since the dopcube is a three-dimensional region, three dimensions is the best place in which to view it. Unfortunately, most of our display devices are 2-dimensional. However, there are excellent graphics workstations with the capability to view 3-dimensional data and manipulate (via translations and linear transformations) this data in real time. We have chosen Evans & Sutherland (E&S) workstations (PS300, 330, 350) for our applications. Figure 1 shows a view of a 21-satellite dopcube taken from an E&S PS330. Time is the vertical axis and the horizontal plane is a cartesian projection of the earth. Hardware dials on the E&S allow the user to view the dopcube from any direction, blow up DOP "clouds", box in clouds to determine their latitude, longitude and time extents, and slice the dopcube in time. John Martillo of The Aerospace Corporation has programmed the E & S to provide these capabilities.

Time Animation of the Dopcube

Mike Werner of The Aerospace Corporation has written a program which performs the following tasks:

1. Slices the dopcube in time to create snapshots at, say, five minute intervals.
2. Overlays time animation of the outages against a world background along with ground tracks and station communication circles.
3. Selects an outage (lat, lon, time) and views the satellite configuration over the (lat, lon) point at that time to determine the nature of the outage.

Figure 2 illustrates a typical display. Highlighted regions indicate degraded coverage, PDOP > 6, at a particular instant in time.

Display of the Dopcube in 2-Dimensions

For more specific analysis of outages, we have several 2-dimensional graphics operators which portray many different aspects of the outages. The graphics are displayed on standard color graphics terminals.

1. Lat-Lon: Shows which regions contain some outage over a given time period. See Figures 3, 6, 7 and 8. The different shades depict different outage durations.
2. Lat-Time and Lon-Time: These projections help the analyst locate when the outages occur for each geographical area of interest. Refer to Figure 4.
3. Duration of outages: Percentage of region vs duration histogram with constellation value. (Figure 5).
4. Area-Time, Area-Lat, Area-Lon: These show how the outages are distributed in time, latitude and longitude, respectively.
5. Overhead View: These show the azimuth and elevation of the satellites in view at an arbitrarily selected user location. These views are very helpful in analyzing degraded coverage regions due to poor geometry, e.g. near co-planar arrangements of SVs. (Figure 9).

SECTION 3. NAVIGATION SERVICE CHARACTERIZATIONS

A complete characterization of the causes of GPS degraded coverage due to poor geometry of the satellites has been achieved for the baseline 21 satellite constellation. At least 4-fold coverage is always possible, i.e., every point on the earth's surface is continuously in view of at least 4 satellites. Figure 3 displays a view of the earth and regions affected by degraded coverage (PDOP > 6). These regions are located at roughly ± 40 latitude with degraded coverage lasting up to 36 minutes consecutive and 65 minutes cumulative over a day, and at roughly ± 60 latitude lasting up to 12 minutes consecutive and 30 minutes cumulative over a day. At the ± 40 latitude, the geometry causes unacceptable errors because the endpoints of the four unit vectors pointing to the satellites becomes coplanar. In these regions of degraded coverage, PDOP rapidly approaches infinity.

The exact times and locations when coverage becomes degraded are shown in Figure 4. The latitude (longitude) vs time plots show, as a function of time, the latitudes (longitudes) where poor coverage occurs given a selected longitude (latitude) band. In terms of the dopcube described in Section 2, these plots are just (partial) projections of the coverage "clouds" onto the latitude-time and longitude-time sides of the dopcube.

Figure 5 gives a general picture of the quality of navigational coverage by plotting a graph of the percent of the earth which experiences degraded coverage for greater than x minutes. A constellation value is also given which gives the probability that a point randomly selected on the earth and in time will have good navigational coverage.

Another concern, in addition to degraded coverage of the baseline 21 satellite constellation, is that in the event of satellite failures, navigational coverage can degenerate fairly rapidly. Figure 6 shows how the quality of navigational coverage can rapidly degrade in the event of a satellite failure in a plane where a spare is not located. The shaded regions indicate different cumulative lengths of time when coverage is degraded or incomplete ($PDOP > 6$).

For these reasons, further work is still being pursued to see if the constellation can be improved. A revision to the current baseline constellation which has been proposed and is under consideration consists of a slight rephasing of the current satellites (phase shifts of less than 17°) to more fully utilize the active spares in the constellation. This provides continuous 5-fold coverage for 98% of the earth, with the remaining 2% of the earth experiencing gaps for at most 5 minutes. It also keeps PDOP bounded by 10.5 at all times and locations on the earth. Figure 7 shows the regions where degraded coverage ($PDOP > 6$) for this constellation can occur. It is, however, only slightly degraded since PDOP never exceeds 10.5. There is also a large increase in reliability because of the improvement in both 5-fold and 6-fold coverage.

Figure 8 shows regions of degraded coverage ($PDOP > 6$) at present using just the current 7 active satellites. This display shows that the current constellation was designed to provide optimal coverage for testing near Yuma, Arizona.

SECTION 4. SUMMARY

The developments described above are evolving analytical/graphical methods used at The Aerospace Corporation to investigate GPS navigation coverage issues. Some of these techniques are being incorporated into on-going U.S. Air Force efforts to provide timely reports on the operational status of GPS navigation service to various classes of GPS users. This 3-D color graphics engineering proto type tool has been installed at the GPS Master Control Station at Falcon AFS in Colorado Springs to support off-line analysis and navigation service status reporting.

Acknowledgments

The authors would like to acknowledge Lt. Col. R. Bowen, USAF, and John Martillo and Mike Werner of The Aerospace Corporation for their assistance in this work. Lt. Col. Bowen has provided direction, encouragement, and support of Aerospace research and development in the area of quantifying the status of the GPS navigational services. Martillo and Werner have provided invaluable software support in the programming of Evans & Sutherland 3-D graphics workstations.

APPENDIX GEOMETRIC FORMULAS FOR GDOP, PDOP, HDOP, AND TDOP

We include GDOP, PDOP, HDOP, AND TDOP formulas where four satellites are used. A formula for calculating HDOP with three satellites is also included.

Let U^i = unit vector pointing from user location to i th satellite.

Let V = volume of tetrahedron formed by the 4 endpoints of unit vectors U^i .

Let b_i = area of i th face of tetrahedron formed by the endpoints of 3 of the unit vectors.

Let V_i = volume of the tetrahedron formed by 3 endpoints of the unit vectors and the origin.

Let a_i = area of i th face of the tetrahedron formed by 3 endpoints of the unit vectors projected onto the plane tangent to the earth's surface.

Then

$$GDOP = \left(\sum_{i=1}^4 b_i^2 + 9 \sum_{i=1}^4 v_i^2 \right)^{1/2} / (3V) ,$$

$$PDOP = \left(\sum_{i=1}^4 b_i^2 \right)^{1/2} / (3V) ,$$

$$TDOP = \left(\sum_{i=1}^4 v_i^2 \right)^{1/2} / V ,$$

$$HDOP = \left(\sum_{i=1}^4 (b_i^2 - a_i^2) \right)^{1/2} / (3V) .$$

If altitude is known and only 3 satellites are used then HDOP is given by

$$HDOP = \left(4\lambda^2(B^2 - A^2) + \sum_{i=1}^3 L_i^2 \right)^{1/2} / (2A) ,$$

where

λ = ratio of altitude error to the 1- σ satellite ranging error,

B = area of the triangle formed by the endpoints of the three unit vectors,

A = area of the triangle formed by the endpoints of the 3 unit vectors projected on to the plane tangent to the earth's surface,

L_i = lengths of the sides of the triangle formed by the endpoints of the unit vectors projected on to the plane tangent to the earth's surface.

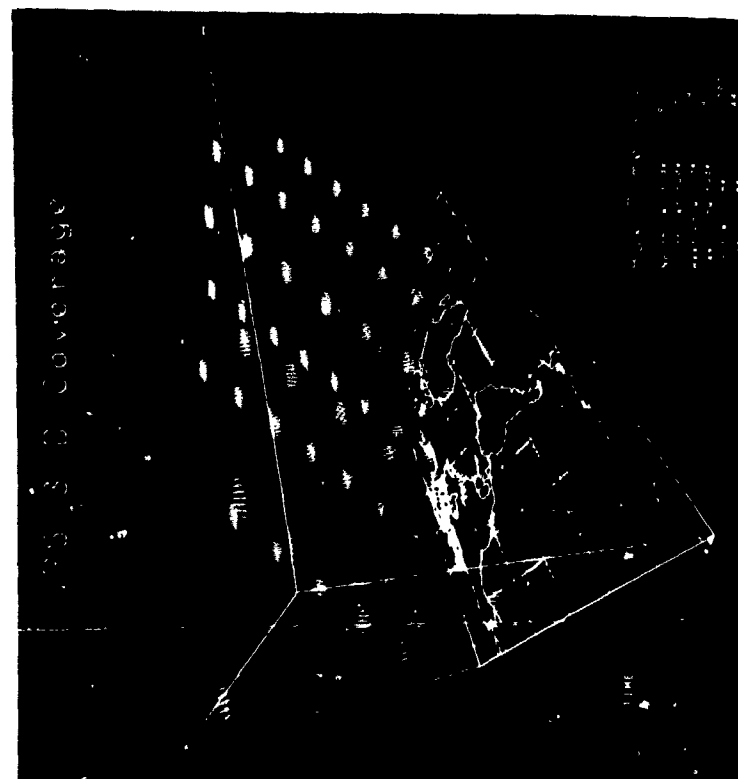


Figure 1. Base line 21 SV Constellation Dopcube.

The horizontal plane containing the world map background is the lat-lon plane. The vertical dimension is the time axis (24 hours). GPS coverage repeats daily with a negative shift in time of approximately 4 minutes. This 3-D Dopcube then completely characterizes the regions of degraded coverage (PDOP > 6) associated with the GPS-21 satellite baseline constellation. Each "outage cloud" is uniquely located in space and time with the thickness of a cloud representing the duration of an outage in this mathematical idealization. On a color graphics workstation, this Dopcube can be viewed from an arbitrary point in space to observe symmetries, etc.

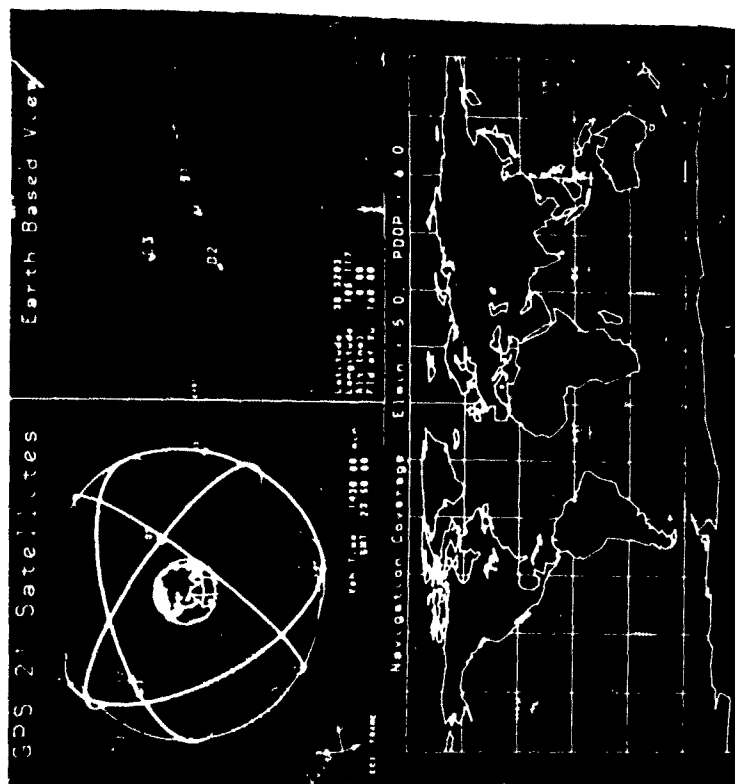


Figure 2. Dynamic Display on EKS PS 390.

This is a snapshot of the time animation of GPS-21 satellite navigation coverage at a fixed instant in time. Satellite sub-points, monitor station locations, terminator (showing those portions of the earth which are sunlit/in darkness), and Greenwich Mean Time can be shown. The shaded areas are regions of degraded coverage at a fixed point in time. This display is frequently used to predict the GPS Navigation service availability to users world-wide at some future date.

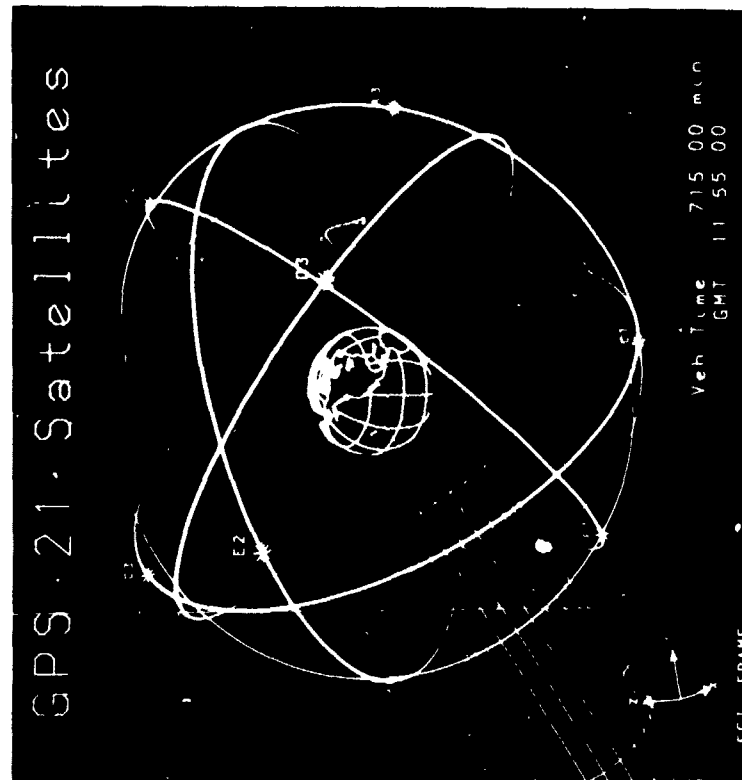


Figure 2a. Overview of the GPS Baseline 21 SV Constellation.

This is a true scale depiction of the GPS Baseline 6-plane 21 satellite system. Earth rotation and orbital motion of the satellites are properly modeled and can be speeded up for convenience of viewing or synchronized for a real-time display. The umbra/penumbra cone can also be shown to display which SVs are in or are approaching a solar eclipse.

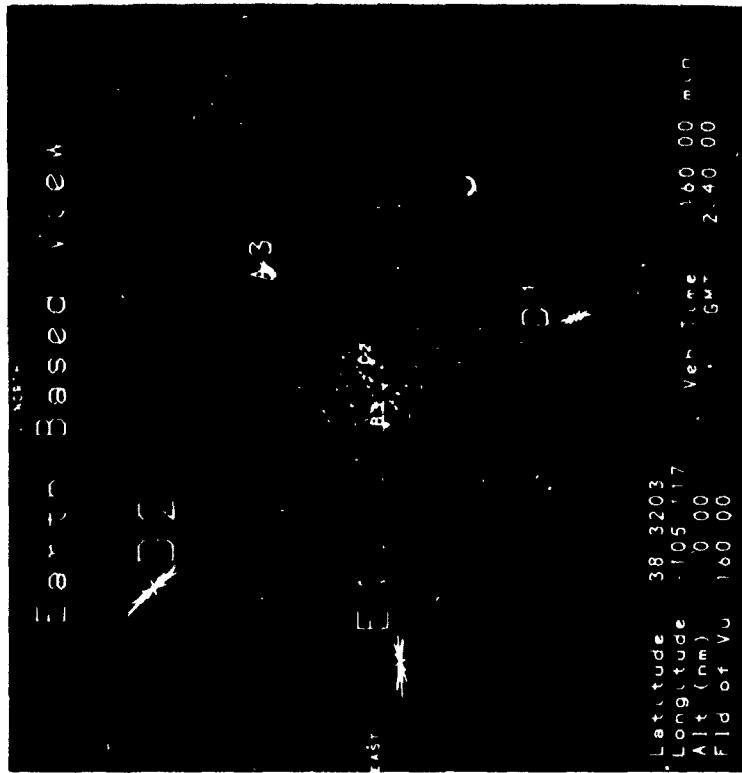


Figure 2b. Overhead View.

This plot shows the azimuth and elevation of all satellites visible to a user at a selected latitude, longitude, and time. This feature is particularly useful in analyzing poor relative geometry of a set of satellites leading to poor (large) DOP values, e.g. near co-planar arrangements, which frequently occur.

DEGRADED COVERAGE 21 SVS (PDOP 6, 3-0=24M)

EPOCH=1985. 7. 1. 0. 0.0 0.0 FROM 0.00 HRS TO 24.00HRS, EL= 5.0
 LATITUDE FROM -90.0 TO 90.0 LONGITUDE FROM -180.0 TO 180.0
 SATELLITES- A1 A2 A3 B1 B2 B3 C1 C2 C3 D1
 D2 D3 E1 E2 E3 F1 F2 F3 A4 E4 C4

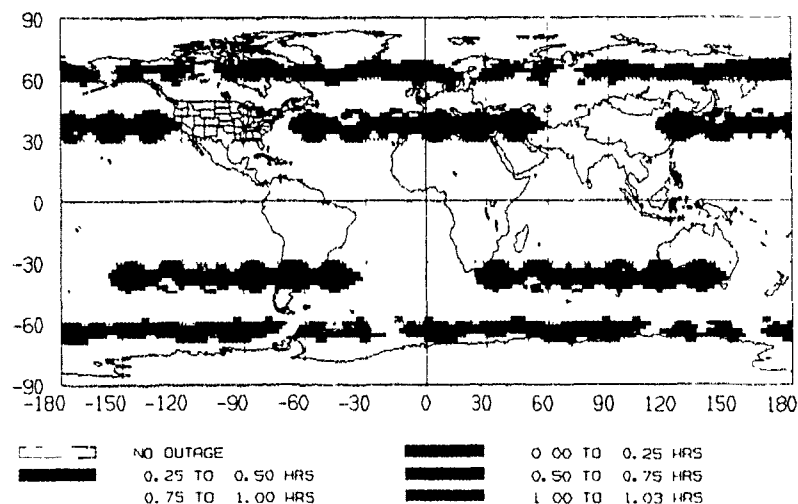


Figure 3. Location of Degraded Coverage Regions.

This display depicts the total accumulated degraded coverage over the span of 24 hours. The longest contiguous span of degraded coverage which probably is of greater interest to navigators can also be displayed. The former is of interest in satellite constellation design and analysis.

DEGRADED COVERAGE 21 SVS (PDOP=6, EL=5 DEG)

EPOCH=1985. 7. 1. 0. 0.0 0.0 FROM 0.00 HRS TO 24.00HRS, EL= 5.0
 LATITUDE FROM 0.0 TO 90.0 LONGITUDE FROM -180.0 TO 180.0
 SATELLITES- A1 A2 A3 B1 B2 B3 C1 C2 C3 D1
 D2 D3 E1 E2 E3 F1 F2 F3 A4 E4 C4

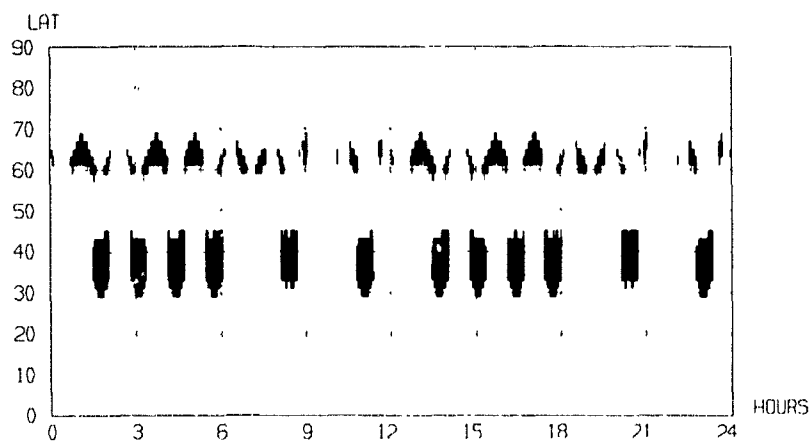


Figure 4a. Time of Occurrence of Degraded Coverage.

Plots 4a and 4b show when as well as where degraded regions of coverage occur for the baseline 21 SV constellation by projecting the regions of degraded coverage onto the latitude-time plane and the longitude time-plane. At any given time a person can determine the regions of the earth likely to have degraded coverage by noting which latitudes and longitudes are shaded.

DEGRADED COVERAGE 21 SVS (PDOP=6, EL=5 DEG)
 EPOCH=1985. 7. 1. 0. 0.0 0.0 FROM 0.00 HRS TO 24.00HRS, EL= 5.0
 LATITUDE FROM -90.0 TO 90.0 LONGITUDE FROM -180.0 TO 0.0
 SATELLITES- A1 A2 A3 B1 B2 B3 C1 C2 C3 D1
 D2 D3 E1 E2 E3 F1 F2 F3 A4 E4 C4

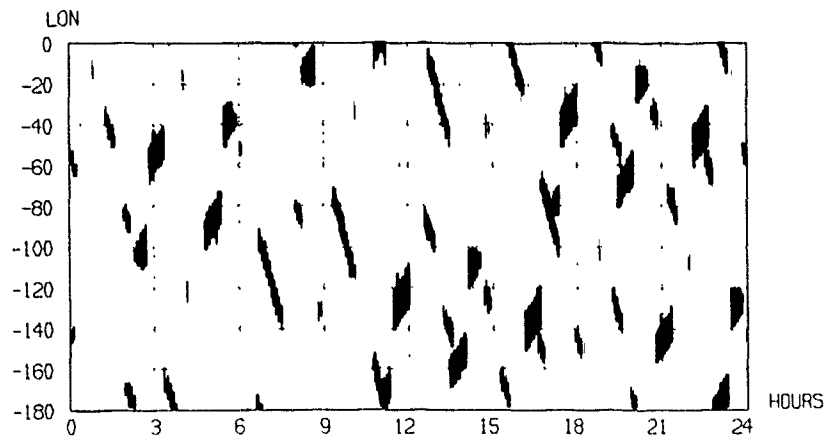
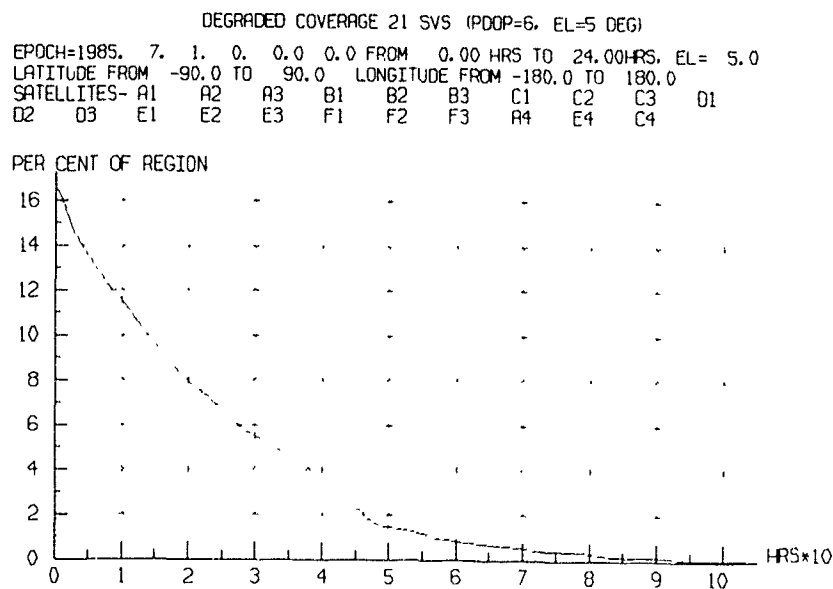


Figure 4b. Time of Occurrence of Degraded Coverage.

Plots 4a and 4b show when as well as where degraded regions of coverage occur for the baseline 21 SV constellation by projecting the regions of degraded coverage onto the latitude-time plane and the longitude time-plane. At any given time a person can determine the regions of the earth likely to have degraded coverage by noting which latitudes and longitudes are shaded.



CV=0.9984

Figure 5. Duration of Degraded Coverage.

This plot shows the general quality of coverage provided by the baseline constellation by plotting the percentage of the earth which has degraded coverage for more than X minutes. The constellation value represents the fraction of time and space where the constellation has coverage.

DEGRADED COVERAGE 20 SV CONSTELLATION (1 FAILED SV)
 EPOCH=1985. 7. 1. 0. 0.0 0.0 FROM 0.00 HRS TO 24.00HRS. CL= 5.0
 LATITUDE FROM -90.0 TO 90.0 LONGITUDE FROM -180.0 TO 180.0
 SATELLITES- A1 A2 A3 B2 B3 C1 C2 C3 D1 D2
 D3 E1 E2 E3 F1 F2 F3 A4 E4 C4

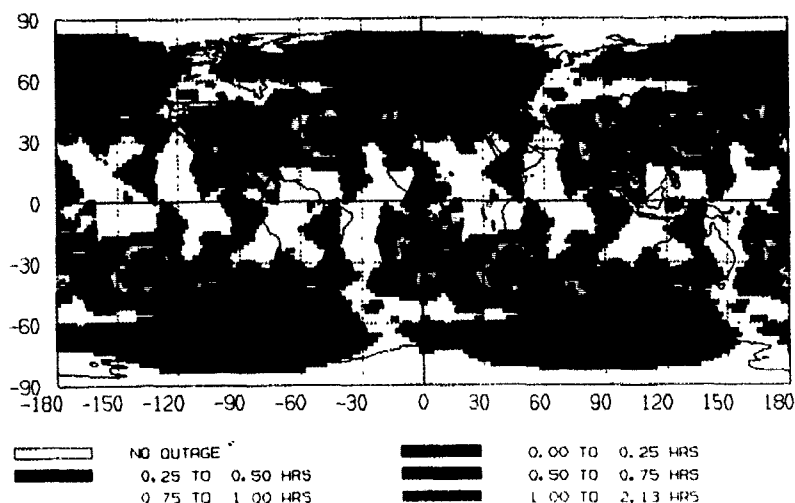


Figure 6. Baseline 21 SV Constellation With One Satellite Failure.

This plot shows the amount of degraded coverage users would experience with just one SV failure. Various regions of the earth get anywhere up to 2.13 hours of degraded coverage per day in this example.

DEGRADED COVERAGE REVISED 21 SVS (PODP=6) CUMULATIVE
 EPOCH=1985. 7. 1. 0. 0.0 0.0 FROM 0.00 HRS TO 24.00HRS. EL= 5.0
 LATITUDE FROM -90.0 TO 90.0 LONGITUDE FROM -180.0 TO 180.0
 SATELLITES- A1 A2 A3 B1 B2 B3 C1 C2 C3 D1
 D2 D3 E1 E2 E3 F1 F2 F3 A4 E4 C4

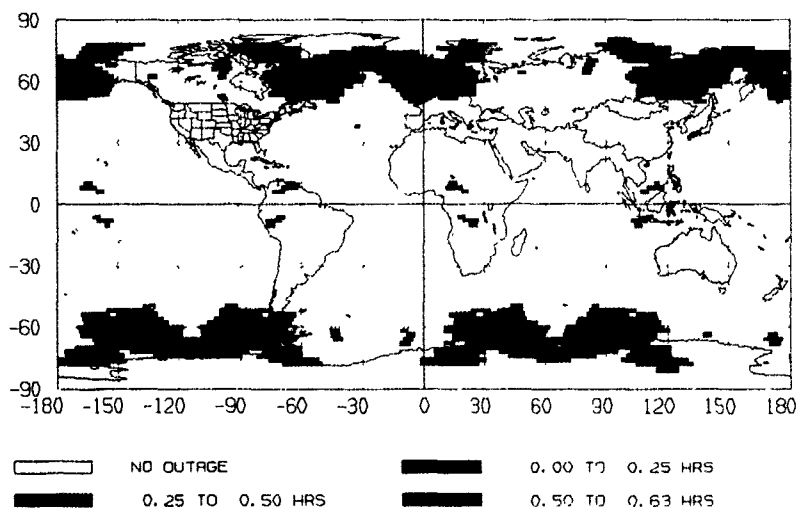


Figure 7. Revised 21 SV Constellation.

The current baseline 6-plane 21 satellite constellation is not optimal. Recent optimization studies show that significant improvements in coverage can be achieved by repositioning satellites in-plane by small phase shifts. This plot shows the coverage provided by a revised 21 SV constellation under consideration.

7 SV NAV COVERAGE (PDOP \leq 6, 3-D=24M)

EPOCH=1988. 5. 1. 0. 0.0 0.0 FROM 0.00 HRS TO 24.00HRS, EL= 5.0
 LATITUDE FROM -90.0 TO 90.0 LONGITUDE FROM -180.0 TO 180.0
 SATELLITES- SVN3 SVN4 SVN6 SVN8 SVN9 SVN10 SVN11

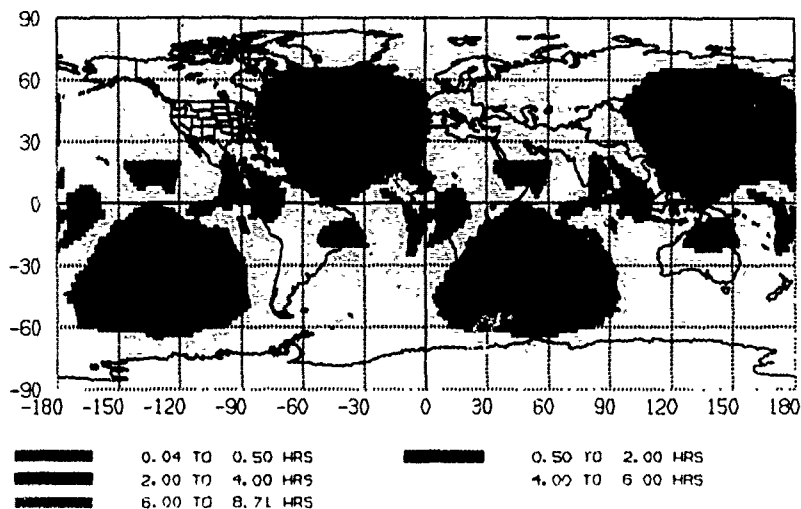


Figure 8. Current 7 Active SVs (PDOP < 6).

This plot summarizes the accumulated total amount of NAV coverage (PDOP < 6) predicted to be provided by the current 7 Block I satellites over a span of 24 hours. GPS coverage is not expected to change significantly from May 1988 through October 1988 when the first Block II satellite is scheduled to be launched.

7 SV NAV COVERAGE (HDOP \leq 25, 2-D=100M)

EPOCH=1988. 5. 1. 0. 0.0 0.0 FROM 0.00 HRS TO 24.00HRS, EL= 5.0
 LATITUDE FROM -90.0 TO 90.0 LONGITUDE FROM -180.0 TO 180.0
 SATELLITES- SVN3 SVN4 SVN6 SVN8 SVN9 SVN10 SVN11

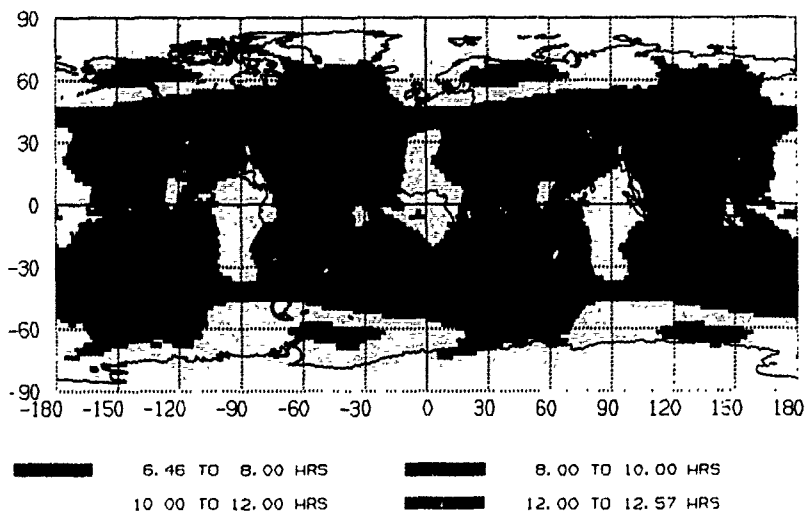


Figure 9. Current 7 Active SVs (HDOP < 25)

This plot summarized the accumulated total HDOP < 25 NAV coverage over a 24 hour span predicted to be provided by the current 7 Block I satellites from May through October 1988. This HDOP coverage was based on 3 satellites with an assumed altimeter uncertainty of 28 meters. With satellite ranging errors averaging less than 4 meters, this HDOP coverage can be interpreted as where navigation to within 100 meters 2-D is possible.

BLOCK-II 3-D NAV COVERAGE (PDOP=6) 9 SVS

EPOCH=1989, 7, 1, 0, 0.0 0.0 FROM 0.00 HRS TO 24.00 HRS, EL= 5.0
 LATITUDE FROM -90.0 TO 90.0 LONGITUDE FROM -180.0 TO 180.0
 SATELLITES- SVN6 SVN8 SVN9 SVN10 SVN11 SVN13 SVN14 SVN16 SVN17

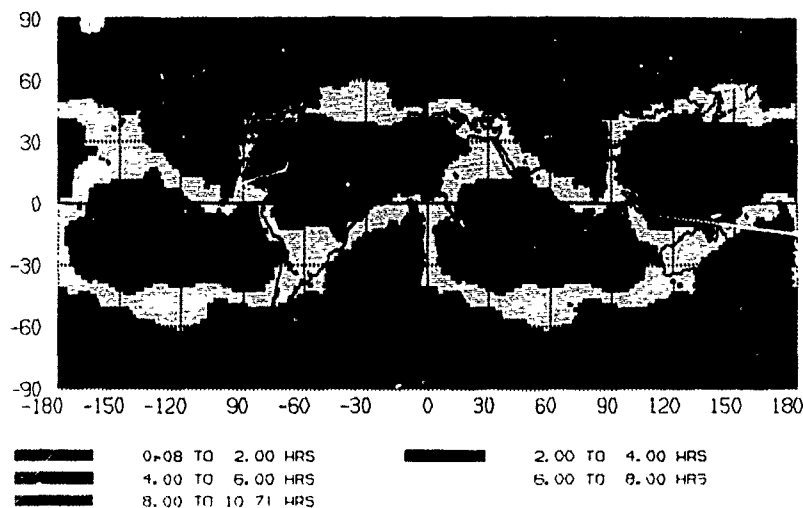


Figure 10. Projected 9 SV Coverage (PDOP < 6).

Summarized in this plot is the total accumulated PDOP < 6 Nav coverage predicted to be provided by GPS in July 1989. This coverage is based on 5 Block I SVs (SVN6, SVN8, SVN9, SVN10, SVN11) and 4 Block II satellites (SVN13, SVN14, SVN16, SVN17) planned to be on orbit on that date. With 1- σ satellite ranging error on the order of 4 meters, the shaded regions can be interpreted as where navigation in 3-D to within 24 meters is achievable.

BLOCK-II 2-D NAV COVERAGE (HDOP=25) 9 SVS

EPOCH=1989, 7, 1, 0, 0.0 0.0 FROM 0.00 HRS TO 24.00 HRS, EL= 5.0
 LATITUDE FROM -90.0 TO 90.0 LONGITUDE FROM -180.0 TO 180.0
 SATELLITES- SVN6 SVN8 SVN9 SVN10 SVN11 SVN13 SVN14 SVN16 SVN17

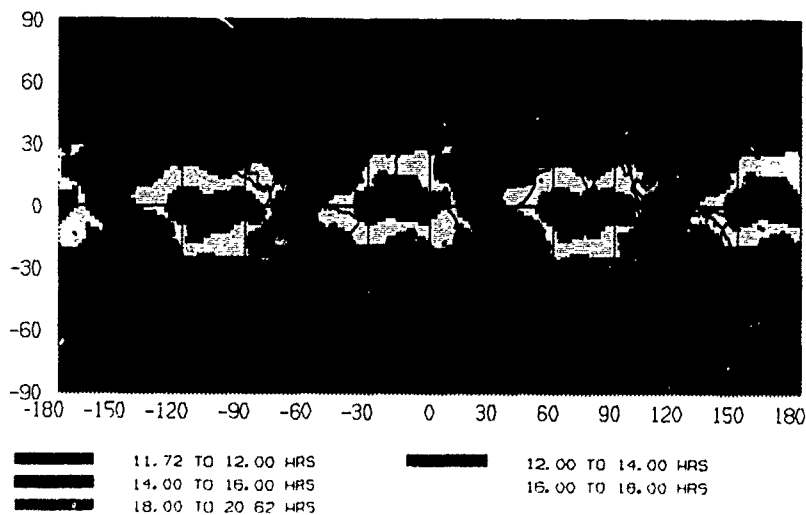


Figure 11. Projected 9 SV Coverage (HDOP < 25)

This plot summarizes the total accumulated HDOP < 25 NAV coverage projected to be provided over 24 hours by 5 Block I satellites and 4 Block II SVs in July 1989. This HDOP coverage is based on 3 SVs and a 28 meter altitude uncertainty. With GPS satellite 1- σ ranging errors averaging less than 4 meters, the coverage depicted can be interpreted as where navigation to within approximately 100 meters is possible.

APPLICATIONS OF DIFFERENTIAL GPS

by

Kjell Hervig
&
Hermod Fjæreide
KONGSBERG NAVIGATION as
P.O.Box 183
N3601 Kongsberg
Norway

SUMMARY

Kongsberg Navigation has developed a system for use of differential GPS for the Norwegian offshore industry.

This article describes the principles of operation and the service offered to customers of Differential GPS. It also describes some of the experiences with the introduction of the service, and prospects foreseen when the system is fully operational.

BACKGROUND

Kongsberg Navigation (KN) has for many years been working with navigation systems and services covering various requirements and applications. Recent involvement has been directed towards the oil exploration activity in the North Sea and in the rest of Norwegian waters. Because GPS originally did not give these users the accuracy they need, Kongsberg Navigation therefore decided to develop a differential GPS system.

In the middle of 1983 KN had already developed its first generation GPS receiver. Based on this know-how, and with the support of several oil companies the project was launched with the goal of having an operating differential system in 1986. The system was named DiffStar, and was after a set of acceptance tests, ready for commercial use Summer 1986. Today two reference stations cover the Norwegian continental shelf and the North Sea. One station is located at Andøya covering the northern area, the other at Askøy covering the southern part and the North Sea. These two stations are transmitting corrections in the 300 kHz band, and have a range of approximately 1000 km over the sea. See figure 1.

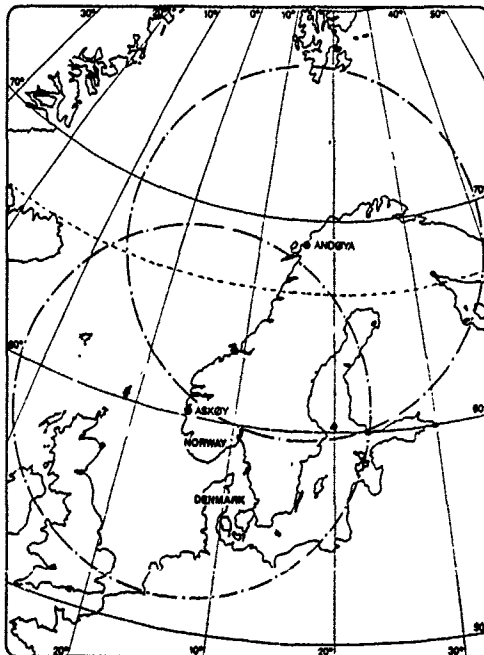


Figure 1 - The DiffStar Coverage by Summer 1987

1. DIFFSTAR - A PSEUDORANGE SYSTEM

The DiffStar system is based on pseudorange corrections not on position corrections like many others. In a differential GPS system based on pseudorange corrections, the reference station calculates the difference between the measured and the theoretical pseudorange. This difference is the pseudorange correction, transmitted to the mobile users as shown in figure 2. The mobile user then corrects his range measurement with the corrections received from the reference station.

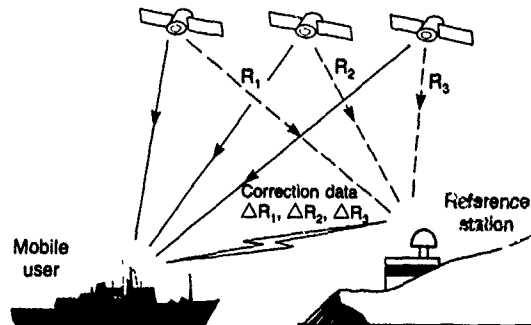


Figure 2 - The Principle of a Differential GPS System based on Pseudorange Correction

RTCM has recommended a standard format for transmitting corrections. The DiffStar system was developed before this standard was determined. The DiffStar system therefore has a slightly different protocol. However, if required, compatibility with RTCM's protocol is easily implemented.

The DiffStar system consists of two parts:

(a) The Reference Station

The reference station is shown in figure 3, and contains the following main functions:

- 2 GPS receivers. To track 8 satellites 2 GPS receivers are normally required. A special 8 channel receiver may, however, be developed if there is a market demand.
- A real time computer reading pseudorange and ephemeris from the GPS receivers, and calculating the pseudorange correction. It is controlling the transmitted data, as well as logging and storing raw data for postprocessing purposes.
- A Cesium clock, used to increase the period of GPS operation. In the future when the system is fully operational this cesium clock can be omitted.
- A printer is available to produce hard copy of data.
- A terminal from which the operator can communicate with the system.
- A modem for remote control of the reference station.
- A transmitter for broadcasting correction data to the users.

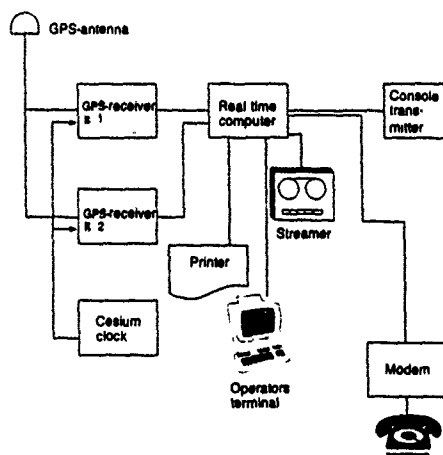


Figure 3 - Block Diagram of the Reference Station

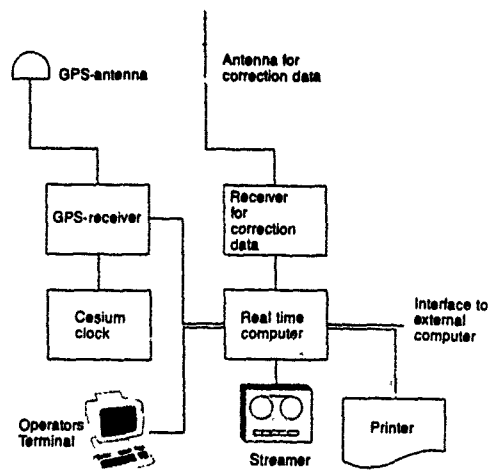


Figure 4 - Block Diagram of the Mobile Station

(b) The Mobile Station

The mobile station is shown in the block diagram in figure 4. The computer reads the pseudorange measurement and ephemeris from the GPS receiver and takes correction data from the 300 kHz radio receiver. Together this is data for computing the differential position.

In order to make postprocessing possible the computer stores all raw data read from the GPS receiver and from the 300 kHz radio. This may be used if the mobile station loses contact with the reference station. Postprocessing also proved valuable in the development phase as this made it possible to collect data in the field with subsequent simulations at the lab in order to debug the software.

The differential correction is updated every 12 seconds. At that time the correction is at least 12 seconds old, and will have to be used for the next 12 seconds until a new message is received. In this way the pseudorange corrections are from 12 to 24 seconds old when used. Therefore, the mobile software has to predict the corrections up to the user time.

(c) The Monitor Station

In a system with several high performance users, a high degree of confidence in the differential GPS system is needed. In order to guarantee the client a specified accuracy a monitor may be set up for the DiffStar system. Due to the fact that the reference station can track all satellites above the horizon a monitor station cannot just be an ordinary mobile station monitoring his own position. All mobile stations in the system can pick its own configuration among the satellites in view at the reference. Therefore, a monitor station in the DiffStar system has to check that the pseudorange corrections obtained from the reference station and the pseudorange correction obtained at the monitor do not differ more than the specified limit. The monitor can then plot the difference in corrections on the reference station and the monitor. If this difference exceeds the limit, the monitor station can dial the system supervisor to give him an alarm. This monitor will also be monitoring that the reference station is alive and transmitting reasonable data.

2. USE OF DIFFSTAR TODAY

Because of the lack of satellites today, the commercial market for DiffStar is limited. In our latitudes we can use the system in two periods each of maximum 4 hours. Obviously, this limits use of the system. Today, DiffStar is mostly utilized as a calibrating system for other radio navigation systems as these users can wait for a period with good geometry.

How is DiffStar used today?

(a) Seismic Explorations

The need for accurate position differs with different kinds of seismic operations. 3D seismic require the highest degree of accuracy. Other seismic operations also appreciate the quality assurance provided by differential GPS.

(b) Surveying

Different surveys like hydrographic investigations, pipe-laying and inspections have used the DiffStar system. Whenever the geometry is good enough, the DiffStar system is used as a calibrating as well as a primary system.

(c) Oil rig moves

There are two phases of an oilrig move that requires accurate positioning. The first phase is the manoeuvre to the intended location. The next phase is to verify the final position. The need for accuracy in this phase is high, but can be obtained by averaging over a period.

These customers are operating on the edge of the limits covered by radio navigation systems, or they are only covered by systems in the 2 MHz range. A 2 MHz system is difficult to use on oilrigs because of the reflection from cranes and other reflecting objects onboard.

In order to use GPS today the users have to synchronize the move of the rig to the coverage of GPS satellites. Up till now Transit translocation has been used to obtain the final position of the rig. To make a good translocation the user has to collect data for 48 hours. In addition there is a post processing phase to obtain the accurate position. Therefore, the advantage of using a real time differential system is obvious. When averaging, a meter accuracy is obtained after a one hour period.

3. RESULTS OF DIFFSTAR

All civilian use of GPS has access to the C/A-code only, and have no opportunity to measure the ionospheric delay. Therefore, the best solution is to use the standard model both in the reference as well as the mobile station. KN has made a set of trials from Norway to UK with a relatively long baseline. Results are good when the geometry is good. When the PDOP is large the inaccuracy of the system increases.

Due to the low number of satellites today the geometry is constantly changing. In most cases it is therefore important to operate with fixed height to reduce the influence of poor geometry.

It is very important that the antenna in a differential system is placed in the best manner to reduce the effect of the multipath. The multipath effect can dramatically reduce the accuracy obtained by the differential technique.

The synchronization of ephemeris in use is also a very important moment in the differential use of GPS. The accuracy of the system will decrease if the mobile and the reference station is using different ephemeris.

The differential technique enables a user to use a satellite of poor quality. Today satellite number 8 (PRN) is operating on a crystal clock. Taking this satellite into the position calculation without any differential corrections will introduce a large error in position. However, when using a differential GPS system, compensation is made for the clock error in the satellite and we obtain acceptable positions.

A typical result from DiffStar is shown in figures 5, 6, 7 and 8. The results demonstrate the improvement obtained by using the differential technique. The non differential solution gives a drift in both latitude and longitude while the differential solution gives a stable position. The drift in position has been removed with the differential technique.

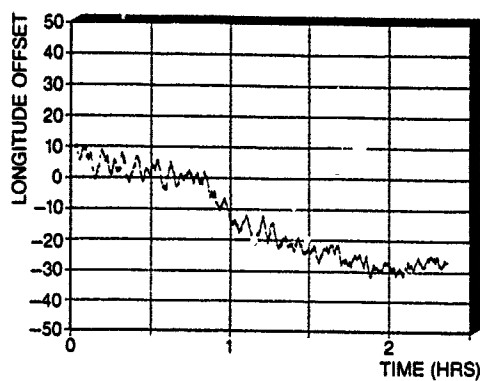


Figure 5 - Non-differential Longitude

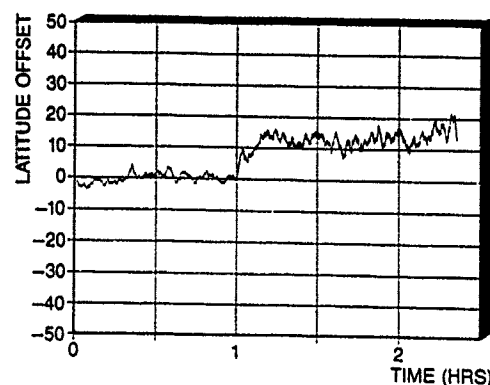


Figure 6 - Non-differential Latitude

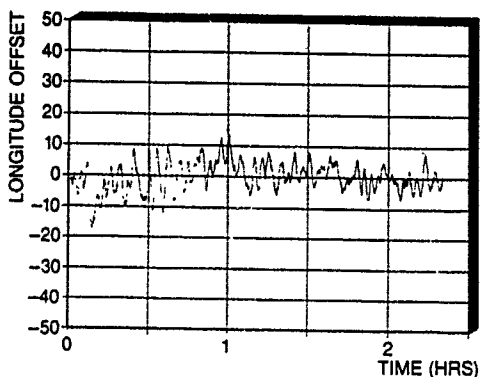


Figure 7 - Differential Longitude

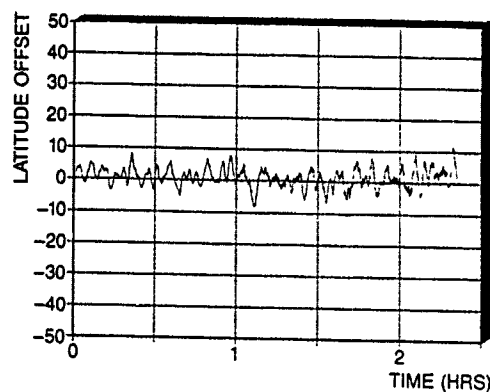


Figure 8 - Differential Latitude

4. FUTURE USERS OF DIFFERENTIAL GPS

The differential system described above has been designed specially to suit the requirements of the Norwegian offshore industry. However, the differential principle has several other potential users. The main problem today is that GPS does not have 24 hours coverage. Several "new" user categories can be included when the system is fully operational.

(a) Military Applications

Within the NATO countries P-code will be available to all military users. For most military operations the accuracy of 17 meter horizontally is adequate. For non-NATO-military users differential GPS may come up as an alternative positioning system. Some applications like the mine operations have an accuracy requirement not met by the P-code. Subsequently, use of differential GPS may be one way to fulfill their needs. In a minesweeping or mine identification operation increased accuracy is directly cost effective. Such operations will not require large longwave transmitter such as the offshore users. Another configuration of the differential principle is to use a small transportable system containing a transmitter covering only the local area. This configuration will require only one geodetic reference point for the transmitting station. Today a local radio navigation system is set up with 3 reference stations.

(b) Harbour Approach

Exact navigation is vital for harbour approach operations in narrow waters. Differential systems may be set up for a local area with only a low power transmitter.

Mobile user equipment may be an instrument handcarried on board the vessel by the pilot. Combined with a powerful mapping display, the trajectory can be predetermined and followed by the helmsman. In such operations the reliability of the system is vital. Use of differential GPS is adding quality assurance to this operation.

(c) Fisheries

Most fishing operations do not require the extra accuracy given by the differential GPS. Cheaper systems like Decca, Loran-C or ordinary GPS give the required accuracy and repeatability. Some operations do, however, require higher accuracies. One of these operations is shell trawling where a high degree of accuracy is required. These operations require big transmitters covering a large area like the system used for the offshore industry. Several users can use the same reference station over a relatively large area.

(d) Dynamic Positioning

The expression Dynamic positioning in this meaning is the system required on board a vessel or platform to control the vessel to follow a predetermined trajectory or to keep a stationary position. This requires accuracies in the region of differential GPS. It is, however, essential to have 24 hours coverage as these users have an extreme requirement to reliability.

(e) Mapping

For hydrographic mapping a transportable differential system will give the required accuracy. In hydrographic mapping operations the data collection phase is the phase where depths and positions are logged for further processing before they are transferred to a data base. Differential GPS will give adequate accuracies in areas which can be covered by relatively small transportable transmitters. Even post processing operations can be used instead of using active transmitters.

For land operations cm accuracy is required in most cases. This is an area where interferometric methods have been used up till now. The challenge of increasing the accuracy of differential GPS uncovers a large market potential.

(f) Air Navigation

The attractive phase of air navigation from a differential GPS point of view is the landing approach. This requires the same accuracy as differential GPS can give.

This is a scenario which can be covered by relatively small transmitters, and with dedicated instruments in the plane. Since GPS is a world wide system this is a very attractive market for the differential technique.

5. CONCLUSION

When GPS becomes fully operational it will revolutionize the world of navigation and positioning. In the world of GPS users there will, no doubt, be room for differential GPS also.

Even if selective availability is not applied here, high performance user will require the extra accuracy which can be provided by the differential technique. One of the most important arguments for the differential technique is however the additional quality assurance and quality control obtained by having an on line monitoring of the GPS system.

AN ANALYSIS OF GPS AS THE SOLE MEANS NAVIGATION SYSTEM IN US NAVY AIRCRAFT

by

Mr George Löwenstein, Naval Air Development Center, Code 409B
Warminster, Pennsylvania 18974-5000

Mr John Phanos, Naval Air Development Center, Code 4013
Warminster, Pennsylvania 18974-5000

Mr Edward C. Rish, SYNETICS Corporation
8233 Old Courthouse Road, Suite 300
Vienna, Virginia 22180-3816
United States

SUMMARY

The Department of Defense (DOD) is developing the Global Positioning System (GPS) to acquire a worldwide navigation capability. This satellite based system provides appropriately equipped users with precision three dimensional position and velocity, and precise time. The current edition of the U.S. Federal Radionavigation Plan, issued in 1984, presents a consolidated federal plan on the management of those Radionavigation systems which are used by both the civilian and military sectors. It states the DOD goal to phase out the use of TACAN, VOR/DME, OMEGA, Loran C and TRANSIT in military platforms and for GPS to become the standard radionavigation system for DOD. This would eliminate all the current sole means air navigation systems (TACAN and VOR/DME) aboard military aircraft. Instrument Flight Rule (IFR) operations within controlled airspace requires an operating sole means air navigation system to be aboard the aircraft. This paper investigates the requirements for GPS certification as a sole means air navigation system in the U.S. National Airspace System (NAS), discusses the implication on GPS User Equipment (UE) hardware and software, describes the actual UE implementation, and discusses approaches for UE integration with flight instruments on Navy aircraft.

1. BACKGROUND

Since the early 1960's both the Navy and the Air Force have actively pursued the idea that navigation and positioning could be performed using radio signals transmitted from space vehicles. The impetus for such a space-based system was the potential for a universal positioning and navigation system which could meet the needs of a broad spectrum of users. In addition, definite cost benefits would accrue by reducing the proliferation of specialized equipments responsive only to particular mission requirements.

The Navy Navigation Satellite System (TRANSIT) became operational in 1964 and was made available to non-military users in 1967. Continuing technology programs investigated the feasibility of a Defense Navigation Satellite System (DNSS). The Navy sponsored TIMATION program involved the development of high stability oscillators, time transfer, and two dimensional navigation. The Air Force concurrently initiated the development of a highly accurate three dimensional navigation system called System 621B.

A key step in the integration of these activities was the memorandum issued by the Deputy Secretary of Defense on 17 April 1973. The memo designated the Air Force as the Executive Service to coalesce the concepts proposed for a DNSS into a comprehensive and cohesive DOD system. A system concept designated NAVSTAR, Global Positioning System (GPS), emerged as a synergistic combination of the best features of the previous navigation satellite concepts and included the Army Position and Navigation requirements.

The GPS satellite-based radio positioning navigation system is designed to provide highly accurate three-dimensional position (to within 16 meters spherical error of probability (SEP) and velocity (0.10 meters per second), and coordinated universal time (to within 500 nanoseconds) to suitably equipped users anywhere on or near (within 500 miles) the earth. Eighteen (18) satellites will be deployed in 10,900 nautical mile circular orbits that are inclined between 45 and 65 degrees and have approximately twelve hour periods. This deployment will provide the satellite coverage necessary for continuous, three dimensional positioning and velocity determination. Each satellite will transmit on two L band frequencies L1 (1575.42 MegaHertz; and L2 (1227.6 MegaHertz). The L1 signal is a composite wave form consisting of a Precise (P code) navigation signal, a Coarse/Acquisition (C/A code) navigation signal, and data such as satellite clock bias information. The L2 signal consists of only the P code.

The GPS receiver normally acquires the C/A code first, then transitions to the P code. Direct P code acquisition can be accomplished by the receiver if it has available to it accurate time and its approximate position. Using the navigation signals from each of four satellites, the receiver measures four independent psuedo-ranges (range to the satellite uncorrected for user clock bias) and psuedo-range rates with respect to the satellites. Measurement of the relative delays between the two L band frequencies allows computation of the ionospheric delay of the signal. The receiver converts this information, in conjunction with the satellite ephemeris data, into three-dimensional position, velocity and time.

2. U.S. RADIONAVIGATION POLICY

To reinforce the goal of creating a cost effective method of reducing the proliferation and overlap of Federally funded radionavigation systems, the United States Congress passed Section 507 of the International Maritime Satellite Communications Act of 1978 requiring the development of a Federal Radionavigation Plan. This resulting joint Department of Defense (DOD) and Department of Transportation (DOT) document established national policy for the future U.S. radionavigation system mix.

The Department of Defense has the responsibility for not only developing, implementing, operating, and maintaining aids to navigation and user equipment required for National defense, but ensuring that military vehicles operating in consonance with civil vehicles have the navigational capabilities required to operate in a safe and expeditious manner. Military support and use of TRANSIT, Loran-C, OMEGA, VOR/DME and TACAN will be phased out, and GPS will become the standard positioning system for DOD. The stated policy also includes the Microwave Landing System (MLS) in the navigation system mix as the precision approach system for aircraft. The total phase out of TACAN on Navy aircraft is contingent upon GPS in conjunction with appropriate communication links, providing a viable replacement for shipboard TACAN, as well as, GPS integrated with other onboard aircraft systems, proving itself acceptable as a sole means navigation system in controlled airspace.

GPS, by definition, will have the potential to become a sole means navigation system for military use. But, for GPS to be so relied upon, adequate navigation by the military user aircraft must be assured in all national and international controlled airspace. DOD/DOT having safety and operational responsibilities, will determine when the GPS integrated aircraft navigation systems are acceptable. The Federal Aviation Administration (FAA) of the DOT is presently developing a National Aviation Standard for civil use of GPS. To obtain insight into the FAA's concept of a navigation capability that can operate in a safe and expeditious manner in controlled airspace, existing documents for the certification of civil air navigation systems were used to draft requirements for GPS development.

3. REQUIREMENTS OF A SOLE MEANS AIR NAV SYSTEM

The basis of the present U.S. National Airspace System (NAS) are the thousands of VOR/DME and TACAN ground facilities throughout the country. The network of VOR/DME & TACAN stations form the fundamental reference grid of the national airways. VOR/DME and TACAN are the only approved systems that can be used for enroute navigation in controlled airspace without the need for any other navigation system, that is, sole means navigation systems. All other navigational sensors including Inertial Navigation Systems (INS) are supplemental systems. The VOR/DME/TACAN station signals provide a positive and unique location beacon and identifier for airway intersections. When a user overflies that signal, it is physically over that geographic location or waypoint.

The introduction of airborne area navigation (RNAV) systems onto the airways has set the stage for the incorporation of GPS. This allows for flights over predetermined tracks without the need to overfly ground-based VOR/DME navigational facilities. Application of RNAV equipment and procedures in the NAS requires that they be compatible with the VOR/DME route structure. Implementation, therefore, means that area navigation devices used must assure proper positioning with respect to the VOR/DME route structure by reference to the geographic locations of the VOR/DME ground facilities. Additionally, the RNAV systems have to provide navigation within the protected airspace of conventional VOR routes, airways, and terminal areas in accordance with established procedures.

For the GPS to fulfill its potential as a Sole Means Air Navigation System, it must satisfy all the requirements of an RNAV system as well as to provide data of equivalent accuracy, stability and verifiability as that of the VOR/DME network. The GPS UE must be designed such that all airborne operations responding to similar Air Traffic Controller instructions will result in similar maneuvering of the aircraft. The system must also provide the pilot with the minimum navigation and guidance functions required for enroute, terminal (including departure and arrival) and approach (including missed approach) phases of flight.

3.1 SYSTEM MECHANIZATION AND ACCURACY

The Minimum Operational Performance Standards (MOPS) for RNAV equipment indicates that "For valid historical, technical and operational reasons, the angular difference between an electronic VOR radial and a geomagnetic radial may be as great as 4-6 degrees. Because all charts used in the NAS show electronic VOR radials for routes and fixes, no significant operational problem results while all aircraft operate in a VOR environment. However, when some aircraft operate in an RNAV (LAT/LONG) environment a navigation system problem becomes apparent due to the angular difference. This problem applies to all RNAV systems used for navigating VOR routes using magnetic courses that have been compensated on the charts for magnetic variations and is due to lack of update of the respective VOR stations." To have the GPS user aircraft totally transparent in the NAS, the UE has to obviate this problem.

While it is possible to project an airway using the location of only one of the terminal VOR stations, the uncertainties in all the correction factors, relating the station to the GPS frame of reference, precludes the use of this "TO-FROM" navigation mode for flying on the airways with acceptable accuracy. Similarly, the chance of misinterpreting a heading command in the magnetic referenced NAS precludes the use of unique DOD airway charts with acceptable risk. If one is to truly replicate a VOR route within a totally GPS aircraft, the UE has to establish the route as perceived by the VOR user. The VOR route, between two ground facilities defining the route, is a great circle path between them (the station signals radiate along great circle paths or radials). Therefore, the UE must be able to compute the great circle path connecting two geographic points (or waypoints) independent of other computations, calibrations or sensor measurements. This mechanization is commonly called a "TO-TO" navigation capability.

The next step is to put limits on the deviations permitted about this direct path to ensure the aircraft remains on the airway. The RNAV MOPS provides the current allowable total error budget (95% probability) for all error sources including Flight Technical Errors (FTE) on airways. This error budget is summarized in Table 1. The cited system errors encompass the entire RNAV suite and its integration aboard the aircraft. The location of the ground stations do not appear in the error budget since they are NAS ground station truth. If we now redefine the system error to include the precision of the ground station position survey (in GPS coordinates), then maintaining these error limits at any distance from the waypoints, defining the route in use, will guarantee the compatibility between GPS area navigation and VOR/DME vector operational environments.

Table 1. 2D RNAV System Accuracy Requirements.
Ref # 6 RTCA DO 187

ERROR TYPE	ENROUTE RANDOM ROUTES		ENROUTE J/V ROUTES		TERMINAL		APPROACH	
	XTK	ATK	XTK	ATK	XTK	ATK	XTK	ATK
SYSTEM (NM)	3.8	3.8	2.8	2.8	1.7	1.7	0.5	0.5
FTE (NM)	1.0	N/A	1.0	N/A	1.0	N/A	0.5	N/A
TOTAL (NM)	4.0	3.8	3.0	2.8	2.0	1.7	0.7	0.5

Verification of the attained system performance for civilian aircraft certification is contained in FAA recommended environmental and performance testing. It is intended to provide a means of determining the overall characteristics of the equipment under conditions representative of those which may be encountered in actual operations. This same intent is embodied in DOD required testing for military equipment which either equals or exceeds the levels recommended by the FAA on both an equipment and total system level.

3.2 SYSTEM AVAILABILITY AND RELIABILITY

For the GPS system to be a sole means air navigation system, it must provide the user with operational availability and reliability as well as accuracy. Additionally, the GPS system must provide the pilot with the assurance that the system performance requirements are being met or else provide him with an unambiguous warning that they are not. The GPS UE continually monitors itself to detect degraded performance. A Figure-Of-Merit (FOM) is developed based upon satellite geometry, the degree on convergence of the receiver and navigation solutions, dependence on external sensor aiding, et cetera. The FOM is continually displayed. Satellite health is checked once per minute. All failure indications are immediately flagged to the operator as well as the FOM falling below preset values.

Waypoints could be reserved for operational verification procedures. For example, if the locations of primary air control radars in the flight path were stored the UE could compute the aircraft range and bearing to the radar installation for comparison with the air controller's radar range and bearing. Such a capability provides a mutual cross check between the onboard GPS system and air traffic control.

The present satellite sparing strategy is to have three active spares in orbit. These spares will be positioned so as to provide optimum coverage for the continental United States. The resulting constellation configuration will provide adequate geometry (Position-Dilution-of-Precision (PDOP) less than 6) for maintaining performance even with a satellite failure.

Aircraft maneuvering can cause a temporary situation where the link between a satellite and the receiver antenna is blocked by the aircraft structure so that only three satellites are in view. The navigation solution can be maintained during this occurrence by receiver aiding. All Navy aircraft installations have been proposed with the UE integrated with a Barometric Altimeter. Knowledge of the aircraft's altitude from an independent source enables the GPS UE to compute position, velocity and time from only three satellites. During normal operations, the UE will continually calibrate the altimeter in preparation for its possible use. Even without this calibration, the altimeter aided three satellites GPS solution has sufficient accuracy to be acceptable.

The following subparagraphs describe the additional operational, control and display functional requirements that were extracted from FAA or Radio Technical Commission for Aeronautics (RTCA) documents (Ref 6 & 7) and were considered significant requirements, all have been designed into the GPS user equipment.

3.3 WAYPOINTS

The receiver should have the capability of storing and recalling, by five-place alphabetic or alphanumeric code, at least 19 waypoints. This number is a low side estimate considering enroute, terminal and approach phases. "In order to minimize pilot workload during non-precision approach, all of the waypoints in the approach (including the initial missed-approach waypoint)" have been included. A second consideration was "that procedures in complex terminals may require the use of up to 6 waypoints in rapid succession for 2-D and up to 10 for 3-D." This led to the requirement that "Means shall be provided, either manually or automatically, to utilize a series of stored waypoints in any selected order."

3.4 FLIGHT INSTRUMENTS

The minimum lateral navigation information required for display to the pilot is the desired track angle, the distance to the active waypoint, and the cross-track deviation from the desired track. In the "TO-FROM" mode (which is mechanized in addition to the "TO-TO" mode in the UE) a continuous display shall be provided to show whether the aircraft is proceeding to or from the active waypoint. In the "TO-TO" mode, a continuous display shall show when the aircraft has passed through the active "TO" waypoint. Maneuvers such as turns to intercept a new course at the active waypoint, must be anticipated when operated in the airspace. The anticipation may be accomplished through computational techniques within the equipment, operational procedures or a combination of both as is mechanized in the GPS UE.

Table 2. ARINC 429 Flight Instrument Port Output Data Words
Ref # 8 ICD-GPS-073

LABEL	PARAMETER	UNITS	RANGE	RESOLUTION	RATE
076	GPS Height Above Reference Ellipsoid	Feet	+131071	1.0	20 Hz
251	Distance to Go	NM	+16383	0.0625	2 Hz
115	Waypoint Bearing (True)	Deg	+180	0.044	20 Hz
117	Vertical Deviation	Feet	+8191	1.0	10 Hz
114	Desired Track	Deg	+180	0.044	2 Hz
116	Cross Track Distance	NM	+128	0.0005	20 Hz
320	Magnetic Heading	Deg	+180	0.044	20 Hz
314	True Heading	Deg	+180	0.044	20 Hz
252	Time To Go	Sec	+131071	1.0	2 Hz
310	Present Position (lat)	Deg	+180	0.000172	1 Hz
311	Present Position (long)	Deg	+180	0.000172	1 Hz
313	Track Angle (true)	Deg	+180	0.044	20 Hz
312	Ground Speed	Knots	+4095	1.0	10 Hz
077	Horizontal/Vertical Deviation	% Full Scale	+128 *	0.8	20 Hz
270	Flight Instrument Discrete Data				1 Hz

*Typical Navy flight instruments display horizontal deviation as a plus or minus two (2) unit (dots on the meter face) needle excursion. Full scale is defined as 128% of the two unit or dot excursion. Flight mode identifiers, contained in the Flight Instrument Discrete Data Word, determine the specific scaling of the Horizontal and Vertical Deviation data. These are listed below in Table 3.

Table 3. Horizontal and Vertical Deviation (Two Dot Excursion)
Ref # 8 ICD-GPS-073

FLIGHT MODE	HORIZONTAL	VERTICAL
ENROUTE	4.00 NM	1000 Feet
TERMINAL	1.25 NM	500 Feet
VERY HIGH ACCURACY	200 Feet	200 Feet
APPROACH	2.344 Deg	0.547 Deg

3.5 DIGITAL DISPLAYS

The equipment shall provide the capability for the display of present position in latitude and longitude with a resolution of 0.1 minutes, and for the observation and amendment of flight plan or other navigation data prior to its utilization. The GPS UE shall also display the distance to the active waypoint and the cross-track deviation from the desired track. The display may be pilot selected and need not be located in the pilot's primary field of view. The cross-track deviation shall be displayed up to a minimum range of 20 NM with a minimum resolution of 0.1 nm up to a cross-track deviation of 9.9 NM and a resolution of 1.0 NM beyond. The distance to the waypoint shall be displayed with a resolution of 0.1 NM or better to a range of 99.9 NM from the waypoint and shall be 1.0 NM or better at ranges greater than 100 NM.

4. NAVY GPS AIRCRAFT SET

4.1 FLIGHT INSTRUMENT INTERFACE

The requirement for flight instrument drive signals was anticipated in the Full Scale Development (FSD) phase where a full set of analog outputs were developed as part of the family of aircraft UE interfaces. The flight instrument capable UE was installed in the Navy's A-6E Test and Evaluation Aircraft and tested for this application. It fed the aircraft's Horizontal Situation Indicator (HSI), which is representative of the flight instruments found aboard Navy aircraft. Concurrent Navy and Air Force integration concept studies indicated that a large proportion of the aircraft inventory already carried a digital-to-analog converter in conjunction with the AN/ARN-118 TACAN. This, coupled with the relatively high weight and power allocations associated with the analog instrument drivers, has prompted an adjustment in the production phase. It was decided to take advantage of the existing interface adaptor unit to drive the analog instruments. The UE will provide flight instrument data digitally to the adaptor unit which will have to be modified to accept it. The choice of electrical format for this UE output was obvious since an ARINC 429 was already required; using this electrical format and imposing ARINC's Electronic Flight Instrument System (EFIS) data format provided the UE with automatic compatibility with future glass instruments.

The Flight instrument data words and refresh rates have been reassessed in light of the FSD test results and have been redefined as summarized above in Table 2.

4.2 CONTROL DISPLAY UNIT

Another outcome of the Navy's integration concept studies and the FSD testing was the need for a Control Display Unit (CDU) with expanded capabilities for consolidation of navigation functions in space limited cockpits and decreased pilot workload, respectively. Additionally, review of the RTCA paper (Ref 7) on software considerations in airborne systems and equipment certification indicates the advisability of partitioning off and isolating the mission and airways navigation functions so as to minimize the need for frequent costly re-certification every time changes are made to the UE. A specification for a GPS CDU has been developed that incorporates and segregates the memory for a certified airways data base from tactical data and the RNAV computations from the general processing. The CDU will also have its own data base input and flight instrument output ports, and will provide a full alphanumeric key set.

The CDU flight instrument port output is identical to its counterpart on the receiver unit. Data processing differs from the receiver in basically two areas. Firstly, most of the manual or external computer activated routines will be automated. An example of which is the automatic sequencing of active waypoints. Secondly, a one thousand (1000) waypoint storage capacity in ARINC 424 format is being provided, significantly increasing the CDU's capacity over the receiver's two hundred (200) waypoint capacity.

To understand the need for a one thousand (1000) waypoint storage capacity, one must keep in mind that the NAS is defined by magnetic radial bearings to VOR/TACAN stations and that all existing sole means navigation systems are based on relative navigation techniques with respect to these VOR/TACANs. A VOR or TACAN station is tuned in by selecting either two or four digits on a control box and your location is established on a specific radial projection from that station. With GPS, you are locating yourself relative to the center of the earth wherein a set of coordinates must be used to define locations upon the earth's surface. A CDU input defining a specific GPS location can require anywhere from 15 to 25 keystrokes. This set of coordinates is defacto set in when you select a TACAN channel or the VOR frequency. The necessities of responding to changing conditions on the airways including air traffic controller redirection precludes enroute insertion of numerous geonetic waypoints as a viable cockpit activity. The one thousand waypoint capacity will allow for the storage of all necessary high and low altitude route structure including airfield departure and approach waypoints.

4.3 CERTIFIED AIRWAYS DATA BASE

The airways data base capability must be provided to assure errors in geodetic coordinates are not inadvertently made. This would be analogous to receiving the ID tone back from a selected TACAN or VOR station, providing a positive feedback to the pilot that the station selected is the one being received. The pilot when operating with a GPS UE must have the same feeling of confidence that he had when operating from existing sole means navigation systems with its positive feedback. The FAA certified

airways data bases are designed to provide the required waypoint data reliably. This is accomplished by maintaining a controlled data source and providing cyclic redundancy checks to guard against data entry.

The ARINC 424 navigation data base is the format currently being used by the commercial aviation industry to generate a verifiable route structure. The data will be loaded via a compatible data loader port and will be mechanized in a manner that provides the FAA required integrity of the data. As mentioned previously, in addition to parity and sum checks, cyclic redundancy checks are built into the data base. Any reformatting of the data can very easily obviate this advantage. Physically, the data base is isolated to prevent back refill that could contaminate the data.

4.4 AIRCRAFT SET COMPLEMENT

4.4.1 RECEIVER PROCESSOR UNIT - In addition to being an RNAV computer, the Aircraft Set is a five-channel, multi-interface equipment that includes an antenna, its antenna electronics, a receiver with integral processor and interface capability to communicate with appropriate aircraft systems, and a control display unit that can also be used to control and display data of various Inertial Navigation Systems (INS). Existing aircraft multi-purpose control and display capability can be used in lieu of the GPS CDU. This external control is exercised via the MIL-STD-1553 multiplex bus and discretes. The Aircraft Set also employs a mountain rack with integral power and signal connectors. Table 4 contains a summary of the physical characteristics of each set element. A relevant description of each unit follows:

Table 4. GPS Aircraft Set Physical Characteristics Ref # 5 Spec # CI-RCVR-3010				
	RECEIVER 3A	ANTENNA ELECTRONICS	FRPA ANTENNA	CONTROL DISPLAY
Dimensions (in)				
Depth	17.1	9.9	4.6	7.0
Width	7.5	4.9	4.6	5.8
Height	7.6	1.9	1.8	6.0
Weight (lbs)	36.0	1.4	0.5	8.0
Power (w)	100.	3.	-	30.

4.4.2 ANTENNA SUBSYSTEM - The antenna subsystem consists of a Fixed Reception Pattern Antenna (FRPA) and an antenna electronics unit.
RECEIVER - The GPS 3A receiver unit performs the receiver and signal processing, the navigation data processing, and interface functions. The receiver accepts and processes the GPS satellite signals received by the antenna subsystem and provides appropriate position, velocity and time data to the aircraft interfaces. RNAV information can be transmitted to the aircraft systems via either the ARINC 429 flight instrument interface or the MIL-STD-1553 multiplex bus interface. Its two hundred (200) waypoint data base can be entered manually using the CDU or electronically via either a RS 422 interface (data loader) or the MIL-STD-1553 multiplex data bus.

4.4.3 CONTROL DISPLAY UNIT - The CDU is the man-machine interface for the GPS set, allowing the operator to control the set and observe set-generated outputs. The CDU can also perform this function for most Navy aircraft inertial navigation systems. As previously discussed, the CDU is an RNAV computer. RNAV information is transmitted to the aircraft systems via the ARINC 424 interface. Waypoints can be manually entered into either the waypoint data base in the receiver or the non-certified portion of the data base residing in the CDU via the keyboard.

5. GPS INTEGRATION OPTIONS

The aircraft set components as just described provide the aircraft integrator with several options for the implementation of an airways capability consistent with existing avionics as well as the national airspace. Table 5 lists the basic configurations available. As noted, not all of the configurations are as desirable as others. The general recommendation is for an RNAV computer, either an existing unit or preferably the GPS CDU with its ARINC 424 data base capability.

The GPS CDU or an existing RNAV computer, in association with the appropriate ARINC 424 data fill device, will provide the recommended configuration. The GPS CDU can be used for this role even when it is not needed or desired to perform the GPS UE control and display function. The CDU then becomes another avionics box and need not occupy prime cockpit real estate. In the reverse role, the CDU if required for control and display of the UE but not for RNAV, can act as a backup to the existing RNAV computer. Other variations of the baseline configurations of Table 5 also exist. In fact, if a more overall mission perspective is taken, the inherent RNAV capabilities can be utilized to the total advantage of the aircraft. For example, the use of all the internal waypoint storage capacity, where a MIL-STD-1553 based mission computer could be using the receiver for tactical navigation and display while relying on the CDU for

Table 5. RNAV Aircraft Aircraft System Options				
OPTION	WAYPOINT DATA BASE UNIT	DATA LOADER INTERFACE	RNAV COMPUTATION UNIT	FLIGHT INSTRUMENTATION INTERFACE*
Recommended	GPS CDU	ARINC 424	GPS CDU	ARINC 429 From CDU
Integration Alternative	Aircraft Specified	Aircraft Specified	Mission or RNAV Computer	Aircraft Specified
Available	GPS RCVR	MIL-STD-1553	GPS RCVR	ARINC 429 From RCVR
Available	GPS RCVR	RS 422	GPS RCVR	ARINC 429 From RCVR
* Via switching unit and/or an interface adapter unit.				

airways navigation and display. Obviously, many possibilities exist. The integration concept on any specific Navy aircraft will be the result of evaluating mission and airways navigation requirements in light of existing onboard avionics and flight safety.

6. CONCLUSION

GPS certification as a sole means air navigation system in the U.S. NAS requires demonstration of the accuracy, availability, reliability and integration of basic guidance data outputs. DOD required testing either equals or exceeds the levels recommended by the FAA on both an equipment and total system basis.

The DOE system including the UE, provides the necessary guidance data well within acceptable accuracies. The UE has been designed to establish within a totally GPS aircraft, the airways as perceived by the VOR/TACAN user. To accomplish this as well as to have the required responsiveness to air traffic control and to minimize pilot workload, a substantial waypoint data base has been included within the UE. The 1000 waypoint data base within the Navy's GPS CDU has been formatted for cyclic redundancy tests to guard against data entry error and is compatible with civil certified airways data bases.

The GPS UE continually monitors itself to detect degraded performance and provides this status to the operator. Altimeter aiding provides the fly wheel to maintain the GPS solution through momentary loss of a satellite signal. With the presently proposed satellite constellation including active spares, the UE can maintain performance within accuracy even with a satellite failure throughout the continental United States.

In summary, the GPS UE when properly configured and integrated with other onboard aircraft systems provides a certifiable sole means air navigation capability for Navy aircraft use in controlled airspace.

7. REFERENCES

1. Aeronautical Radio, Inc., "Area Navigation System Data Base", ARINC Characteristic 424-5, March 18, 1985.
2. Aeronautical Radio, Inc., "Electronic Flight Instruments (EFI)", ARINC Characteristic 725-2, September 5, 1980.
3. Aeronautical Radio, Inc., "Mark 33 Digital Information Transfer Systems (DITS)", ARINC Characteristic 429-10.
4. Collins Government Avionics Division, "Computer Program Development Specification for the GPS Receiver (RCVR 3A) CPCI of the User System Segment", Specification No. CI-RCVR-3010, November 8, 1985.
5. Collins Government Avionics Division, "Prime Item Development Specification of the GPS Receiver (RCVR 3A) of the User System Segment", Specification No. CI-RCVR-3010, November 10, 1985.
6. Radio Technical Commission for Aeronautics, Special Committee 137 (SC-137), "Minimum Operational Performance Standards for Airborne Area Navigation Equipment Using Multi-Sensor Inputs", Document No. RTCA/DO-187, November, 1984.
7. Radio Technical Commission for Aeronautics, Special Committee 152 (SC-152), "Software Consideration in Airborne Systems and Equipment Certification", Document No. RTCA/DO-178A, March, 1985.
8. U.S. Department of Defense, "GPS User Equipment - Digital Flight Instruments (ARINC 429) Interface", Interface Control Document ICD-GPS-073, GPS-86-12320-036, December 6, 1985.
9. U.S. Department of Defense, "GPS User Equipment - MIL-STD-1553 Multiplex Bus Interface", Interface Control Document ICD-GPS-059, GPS-86-12320-037, November 30, 1985.
10. U.S. Department of Defense, "GPS/INS Control Display Unit (CDU) Specification, December 19, 1985.
11. U.S. Department of Defense/Department of Transportation, "Federal Radionavigation Plan, 1984", DOD-4650.4/DOT-TSC-RSPA-84-8, 1984.
12. U.S. Department of Transportation, Federal Aviation Administration, "Approval of Area Navigation Systems for Use in the U.S. National Airspace System", Advisory Circular AC90-45A, February 21, 1975.
13. U.S. Department of Transportation, Transportation Systems Center, "NAVSTAR GPS Simulation and Analysis Program", DOT-TSC-RSPA-83-11, Final Report, December, 1983.

THE DETERMINATION OF PDOP (POSITION DILUTION OF PRECISION) IN GPS

by

Alan H. Phillips, Engineering Consultant
3 Honey Drive
Syosset, NY 11791
United States

SUMMARY

PDOP (Position Dilution of Precision) is defined, and equations are given to calculate it. Equations are given for 3-dimensional GPS fixes, 2-dimensional GPS fixes, 2-dimensional hyperbolic fixes (Loran), and 2-dimensional range-range fixes. A method is given for geometrical determination of PDOP. The method gives an insight, which is lacking in the purely mathematical determination. Practical examples are given, and the results of the geometric determination are shown to agree with the purely mathematical determination. An equation is given for a dilution factor which applies to determination of velocity; it is not the same as PDOP.

DEFINITION OF PDOP

PDOP relates error in GPS position to error in pseudo-range to the satellite. (A pseudo-range measurement is made by measuring the time of arrival of a signal from the satellite with respect to a timing source located at the receiver. It is termed pseudo-range, because the timing source has an unknown offset. This offset is determined, and its effect eliminated by measuring the arrival times of four satellites. For further details of the GPS position fix process see reference 1.)

$$\sigma_p = (\text{PDOP}) \cdot \sigma_r$$

σ_p is the rms radial error in position (meters)

σ_r is the rms error in each pseudo-range measurement (meters)

Errors can be in the arrival time measurement, or can be due to uncertainties in the knowledge of satellite position, or due to errors in the timing sources of the individual satellites. The rms range error due to all of these sources is termed User Equivalent Range Error (UERE). The offset in the receiver's timing source does not contribute to the pseudo-range error.

For PDOP to have meaning, the errors must be random, have zero bias, and have equal rms value for all satellites.

CALCULATION OF PDOP

PDOP is calculated by the following formula, in the case of GPS 3-dimensional position:

$$\text{PDOP} = \left\{ \text{TRACE}_3 \left[G^T G \right]^{-1} \right\}^{1/2}$$

G is the matrix relating changes in measured pseudo range to changes in position and time reference

i.e.

$$\begin{bmatrix} \delta r_1 \\ \delta r_2 \\ \delta r_3 \\ \delta r_4 \end{bmatrix} = \begin{bmatrix} G \end{bmatrix} \begin{bmatrix} \delta x \\ \delta y \\ \delta z \\ \delta t \end{bmatrix}$$

TRACE_3 denotes the sum of the first 3 diagonal elements of $G^T G^{-1}$

$\delta r_1, \delta r_2, \delta r_3, \delta r_4$ are changes in pseudo-ranges to the 4 satellites.

δt is the change in timing of the on-board time reference

$\delta x, \delta y, \delta z$ are changes in receiver position in a local coordinate frame.
(x east, y north, z up)

The origin of the local coordinated frame is a fixed point on earth at the present position of the receiver.

The G matrix, first 3 columns, has elements which are negatives of the x, y, and z components of unit vectors to the satellites; the 4th column elements are all (-1)

GEOMETRIC INTERPRETATION OF PDOP

PDOP is determined geometrically as follows:

- 1) Four unit-vectors point toward the satellites (see Fig. 1). Connect the ends

of these vectors with 6 line segments, forming a tetrahedron.
 2) PDOP is the r.s.s. (square root of the sum of the squares) of the areas of the 4 faces of the tetrahedron, divided by its volume. Expressed somewhat differently, it is the r.s.s. of the reciprocals of the 4 altitudes of the tetrahedron.

Proof of the above is given in reference 2. The construction is 3-dimensional, but it is not difficult, as will be seen in the following example.

At a particular time and location, Satellites 1, 10, 15, and 18 will have the following elevations and azimuths:

Satellite	Elevation	Azimuth
1	37.6°	130.9°
10	65.1°	326.6°
15	22.8°	289.5°
18	40.6°	39.8°

The corresponding components of the unit vectors to the satellites will be as follows:

Satellite	x (east)	y (north)	z (up)
1	.5987	-.5185	.6105
10	-.2318	.3522	.9068
15	-.8685	.3080	.3883
18	.4865	.5831	.6506

From the Pythagorean Theorem, the distances between the ends of the unit vectors are:

10 - 1	1.2392
15 - 1	1.6986
18 - 1	1.1080
15 - 10	.8223
18 - 10	.7968
18 - 15	1.4073

Knowing these distances, a tetrahedron can be constructed. The individual faces are constructed as shown in fig. 2, and the figure is then cut out and folded to make a tetrahedron. The 4 altitudes are then measured. The results of the measurements are

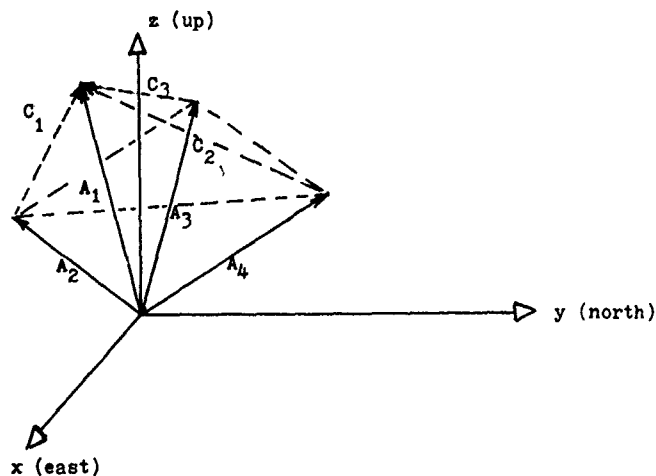
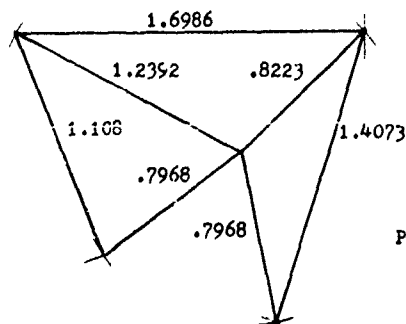


Fig. 1 - Unit vectors to satellites



Altitudes of
Tetrahedron

$$h_1 = 1.06$$

$$h_{10} = .39$$

$$h_{15} = .70$$

$$h_{18} = .63$$

$$\text{PDOP} = \left[\frac{1}{h_1^2} + \frac{1}{h_{10}^2} + \frac{1}{h_{15}^2} + \frac{1}{h_{18}^2} \right]^{1/2}$$

$$= 3.46$$

Fig. 2 - "Cut and Fold" tetrahedron for PDOP determination

given in the figure, and PDOP is determined using the formula in the figure. The value (3.46) agrees well with the value determined from equation (1) (3.44). The discrepancy is due to small errors in constructing the tetrahedron, and in measuring the altitudes. There are no approximations associated with the geometric expression.

If the angle 18 - 10 - 18 is small, one can recognize that PDOP will be large (poor) without cutting out the figure, since this leads to a tetrahedron having a small volume.

A consequence of the above construction is that if the ends of the unit vectors are co-planar (a not unusual circumstance) PDOP becomes infinite, and a position is not obtainable. For this reason, the final GPS constellation will be such that there will almost always be at least 5 satellites in view anywhere on earth; giving an alternate choice of the 4 satellites to be utilized.

GPS 2-DIMENSIONAL POSITION ACCURACY

Unfortunately there is no analogous neat geometrical construction for the horizontal component of PDOP, HDOP. This quantity is of importance to surface vehicles.

If altitude is accurately known, however, a 2-dimensional position is obtainable from pseudo-range measurements on 3 satellites, and this 3-satellite PDOP (HDOP) can be obtained by a geometrical construction.

In this case:

$$PDOP = HDOP = \{ \text{TRACE}_2 [G^T G]^{-1} \}^{1/2}$$

where: G relates changes in measured pseudo-range to changes in 2-dimensional position and time.

i.e.:

$$\begin{bmatrix} \delta x_1 \\ \delta x_2 \\ \delta x_3 \end{bmatrix} = \begin{bmatrix} G \end{bmatrix} \begin{bmatrix} \delta x \\ \delta y \\ \delta t \end{bmatrix}$$

This 2-dimensional G matrix is the previous 3-dimensional G matrix with the third column deleted, and the row deleted corresponding to the unused satellite.

Figure 3 shows the construction. Draw the horizontal components of 3 unit vectors pointing toward the satellites. Connect the ends of these vectors with 3 line segments, forming a triangle. PDOP (HDOP) is the r.s.s. of the lengths of the sides of the triangle, divided by its area. Expressed differently, it is the r.s.s. of the reciprocals of the 3 altitudes of the triangle.

Using the previous case; satellites 1, 15, and 18: From the Pythagorean Theorem, the distances between the ends of the horizontal projections of the unit vectors to 1, 15, and 18 are:

1	15	1.684
15	18	1.382
18	1	1.107

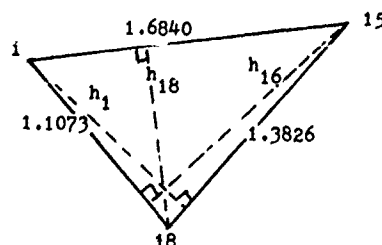
Knowing these distances, the triangle in Fig. 3 is constructed, and the altitudes are determined. HDOP is calculated from the formula in the figure. The result agrees well with calculations from matrix theory (1.60).

Knowing the altitude accurately for a surface vessel requires knowing accurately the geoidal height of the ocean in the vicinity of the vessel. An error in knowledge of geoidal height translates into a bias error in position.

PDOP becomes infinite when the ends of the 3 vector components are colinear.

GPS VELOCITY ACCURACY

Three-dimensional velocity can be determined from GPS signals by measuring range rates to three satellites. Range rate measurements are determined from phase measurements on the microwave carrier, and are inherently very accurate. The dilution of precision for velocity measurements is not PDOP. It will be termed VELDOP in this description.



$$HDOP = \left[\frac{1}{h_1^2} + \frac{1}{h_{18}^2} + \frac{1}{h_{15}^2} \right]^{1/2}$$

Fig. 3 - 2-Dimensional PDOP Determination

$$\sigma_v = (\text{VELDOP}) \sigma_r$$

σ_v is the rms velocity error in meters/sec.

σ_r is the rms range rate error in meters/sec. on each satellite signal

$$\text{VELDOP} = \left\{ \text{TRACE}(G^T G)^{-1} \right\}^{1/2}$$

G is the matrix relating range rates to velocity

$$\begin{bmatrix} \dot{r}_1 \\ \dot{r}_2 \\ \dot{r}_3 \end{bmatrix} = \begin{bmatrix} & & \\ & G & \\ & & \end{bmatrix} \begin{bmatrix} \dot{x} \\ \dot{y} \\ \dot{z} \end{bmatrix}$$

The G matrix elements are the x,y, and z components of unit vectors to the 3 satellites. VELDOP can be determined geometrically by constructing the tetrahedron of the 3 unit vectors to the satellites. VELDOP is the rms of the reciprocals of the altitudes of this tetrahedron.

LORAN POSITION ACCURACY

The above construction can also be used for 2-dimensional hyperbolic systems having all transmitters on the surface of the earth, such as Loran, and Decca. In this case the vectors are unit length and in the horizontal plane. PDOP become infinite when the ends of these vectors are colinear, which occurs on baseline extensions.

RANGE-RANGE POSITION ACCURACY

An analogous construction can be used for 2-dimensional range-range systems; both transmitters on the surface of the earth. Examples are Autotape, and Raydist. In this case, unit vectors are drawn to the two stations, and a triangle is drawn by connecting the ends. PDOP is the r.s.s. of the reciprocals of the altitudes to two of the sides (the unit vectors). Figure 4 shows the construction. In this case it can be geometrically shown that $\text{PDOP} = \sqrt{2}/\sin \theta$.

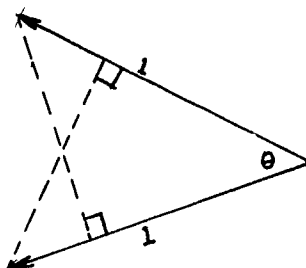


Fig. 4 - 2-Dimensional range-range PDOP determination

REFERENCES

1. Milliken and Zoller, "Principle of Operation of NAVSTAR and System Characteristics", Navigation, Summer 1978.
2. Phillips, "Geometrical Determination of PDOP", Navigation, Winter 1984-85.

PART III

Optical Gyroscope Guidance and Control Systems

RING LASER GYRO PRINCIPLES AND TECHNIQUES

by

Graham J. Martin, PhD
Member Technical Staff
Optical Technology
Litton Guidance & Control
5500 Canoga Avenue
Woodland Hills, CA 91367
United States

SUMMARY

The ring laser gyroscope (RLG) represents a departure from the mechanically based inertial rotation sensor used since the turn of the century. Conventional rotation sensors operate on mechanical principles derived by Euler over two hundred years ago and make use of the torques produced by a spinning rotor. Since their use in the first crude gyroscopes manufactured for gun platforms at sea before the first World War, these systems have been highly refined to give the compact, accurate and relatively cost effective systems of today. In contrast the RLG works on purely optical principles and involves no moving parts. It offers true strapdown capability and has no known sensitivity to high g fields. Other attractive features include almost instant turn-on performance and a very linear scale factor over a very large dynamic range. Development of the RLG started in the early 1960's with the first flight tests being performed in the early 1970's. Commercial RLG-based inertial guidance systems have been available for about a decade, however the device has perhaps yet to reach its full potential.

The ring laser gyro is one configuration in a class of optical gyroscopes, all of which are based on a relativistic principle known as the Sagnac Effect. As early as 1897 Oliver Lodge first put forward the possibility that rotation could be sensed inertially using circulating light beams although it was left to Georges Marc Marie Sagnac to actually derive the basis mathematically and carry out preliminary experiments in 1913. However the size of the Sagnac Effect was regarded as too small to be of practical use for inertial navigation at that time as is graphically illustrated by the famous experiment of A. A. Michelson. He laid out a ring of evacuated pipe almost a mile in perimeter in the farmlands of Illinois in 1925 and by sending light beams in opposite directions around this loop was barely able to detect the rotation of the Earth. The advent of the laser in 1959 and the renewed interest in optical resonant cavities provided the impetus needed for the first practical optical gyroscope designs.

This article on the ring laser gyroscope describes briefly the Sagnac Effect and outlines how it may be used to build optical rotation sensors. The RLG is placed in perspective with other forms of optical gyroscope and the basic layout is described. Details are given of the more serious error sources, such as frequency locking, and the most common means for minimizing them. In particular the lockin reduction scheme known as mechanical dither is explained. Finally an alternative form of the RLG known as the multioscillator is outlined as a possible configuration for systems where the mechanical dither approach may have shortcomings.

OVERVIEW OF OPTICAL GYROSCOPES

The Sagnac Effect

The Sagnac Effect refers to two beams of light propagating in opposite directions around a closed loop such as a ring of mirrors or fiber optic loop. If the loop is not rotating in inertial space, while the light beams are propagating inside, then the times taken for the beams to complete the loop are identical. However if the loop does rotate then the two propagation times are different; the velocity of the light is the same but the beam travelling with the direction of rotation sees a longer light path than the beam going against the rotation.

The diagram in Figure 1 depicts an idealized circular light path of radius R along which the counterpropagating beams may propagate. The light is injected into the path at point A and while in transit the loop rotates with angular rate ω so that A is displaced to the position A'. As the light exits at A' the beam moving against the rotation will have travelled a shorter distance than the other beam. This path difference is given by

$$\Delta L = 4A\omega/c$$

(1)

If the exiting beams are made to interfere on a screen at S then the pattern of bright and dark fringes created will be shifted according to the path difference induced by the Sagnac Effect. Figure 1 shows qualitatively the relative fringe pattern position for a stationary and a rotating loop. The position of the fringe pattern is defined by the rotation rate of the loop.

The path difference given in Equation (1) by this simplified model is consistent with that derived more rigorously using Einstein's General Theory of Relativity and a more generalized light path; the Sagnac Effect is a relativistic phenomenon.

III-2

The technique shown in Figure 1 of allowing the exiting beams to interfere to create a fringe pattern, was used by Sagnac and later by Michelson to demonstrate the Effect and such a configuration is known as a Sagnac interferometer. The magnitude of the fringe movement depends on the enclosed area of the loop and is very small. Thus Michelson,

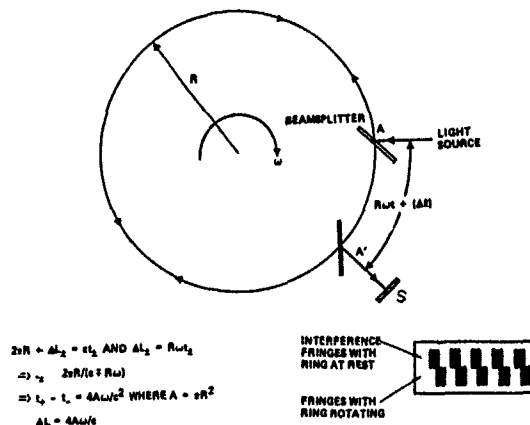


FIGURE 1

with an enclosed area of about a sixteenth of a square mile, only observed a shift in fringe position of about two tenths of a fringe spacing due to Earth's rotation (equivalent to about 10 deg/hour in the plane of Michelson's loop in Illinois). Direct use of a Sagnac interferometer as a practical means of making a gyroscope has only been possible in the last decade with the use of very low loss optical fibers and sophisticated electronics for signal processing. The inherent insensitivity of the Sagnac Effect is overcome by using many thousands of optical loops (typically several kilometers of fiber) thus multiplying the path differential between the counterpropagating light beams. The ring laser gyroscope offers an alternative approach to this. As described later this device makes use of a hand-palm-sized resonant ring cavity with an enclosed gaseous gain medium to greatly enhance the basic Sagnac Effect.

Principles of Resonant-Cavity Gyros

The small path-length difference created by the Sagnac Effect can be converted into a frequency difference by the use of an optically resonant cavity. Linear optical cavities consisting of two mirrors arranged face-to-face so that light can bounce back and forth almost indefinitely have been commonplace since the discovery of the laser. A ring cavity, usually a triangle or square, can be similarly formed with three or four mirrors and creates the light loop necessary for the Sagnac Effect to take place. Counterpropagating beams will now propagate indefinitely (in the absence of cavity losses) in the loop in contrast to the single pass made in the Sagnac interferometer. The cavity becomes resonant when the light path in the ring is equal to an exact multiple of the wavelength of light travelling therein. Thus if the physical perimeter of the ring is P , any wavelength of light which obeys the condition

$$m\lambda = P = mc/f \quad (2)$$

where m is an integer, c is the velocity of light in the cavity and f is the light frequency, will resonate. The integer m is known as the mode number and for typical RLG-sized cavities of around 20cm is of the order of a third of a million. There are many wavelengths of visible light that will resonate, limits being placed in general by the frequency range over which the cavity mirrors can give high reflectivities.

An important parameter for a given cavity is the frequency difference between consecutive wavelengths, say m and $m+1$ that resonate, rather than the absolute frequencies. This quantity is called the free spectral range (or longitudinal mode spacing). It is easily derived from the cavity closure condition.

$$\text{For mode } m \quad m\lambda = P = mc/f_m \quad \Rightarrow \quad f_m = mc/P$$

$$\text{For mode } m+1 \quad (m+1)\lambda = P = (m+1)c/f_{m+1} \quad \Rightarrow \quad f_{m+1} = (m+1)c/P \quad (3)$$

$$f_{m+1} - f_m = c/P = \text{free spectral range}$$

A typical free spectral range for an RLG is about one gigahertz.

Any apparent change in the path length such as that created by the Sagnac Effect will alter the wavelength and frequency of light resonating in the cavity. When the relationship between cavity length and resonant frequency from Equation 3 is substituted into the Sagnac path shift the following result is obtained.

$$\text{Output (Hertz)} = \frac{4 \cdot A}{P \lambda} \text{ Input rate (rads/sec)} \quad (4)$$

where A is the enclosed area of the ring. Thus in the presence of inertial rotation the resonant frequencies for the light beams propagating in opposite directions around the ring will be different. The very short wavelengths of light compared to a typical cavity length of say 20cm provides a very large scaling factor and minute fringe shifts are converted into easily discernible frequency differences. For instance the Earth's rotation rate in a typical 20cm cavity gives about a 4 Hz frequency difference. Michelson's one-mile ring would have given about 31kHz if, in principle, used as a resonant cavity.

The use of a resonant cavity alone to detect the Sagnac Effect is known as the passive cavity optical gyroscope. Light from a broad-band external source is directed into the cavity in each direction. The cavity filters the light and transmits a narrow band, the peak frequency of which depends on the direction of propagation in the presence of any Sagnac Effect. Currently some efforts are underway to investigate this technique as a basis for a compact guidance system, but most interest in the passive cavity gyro is centered around scientific use for astronomical observations of rotating frames.

The resonance linewidth of a cavity is proportional to the light losses within the cavity. The cavity losses are an important parameter for resonant cavity gyroscopes and are often described in terms of the cavity finesse which is defined as the ratio of the free spectral range (fsr) to the cavity transmission linewidth. In practice the fractional cavity losses are approximately equal to 2π divided by the finesse.

$$\text{Finesse} = \frac{2\pi}{\text{loss}} = \text{fsr/bandwidth} \quad (5)$$

Current mirror technology will give cavities with finesse of several tens of thousands or linewidths down to about 10kHz. Even modest navigation requirements mean measuring the center of such a linewidth with an accuracy of a fraction of a Hertz and, although practically possible, the manufacturers of optical inertial navigation equipment favor a method for circumventing this difficulty, namely the ring laser gyroscope.

The RLG consists of a resonant cavity containing an active gain medium, usually in the form of a gaseous discharged electrically. This medium not only provides an internal light source for the Sagnac Effect operating at exactly the cavity resonant frequency, but by means of stimulated emission, replaces photons lost from the cavity with duplicates of the same frequency and phase. Thus effective linewidths are reduced by many orders of magnitude over the passive cavity and the very precise beat between counterpropagating beams is easily observed when small fractions of each beam are allowed to exit the cavity through partially transmitting mirrors and combine to form a fringe pattern. The position of this pattern is dependent on the angular position of the gyro frame in contrast to the Sagnac interferometer where the position depends on the angular rate. A light detector is used to count the passing of bright and dark fringes and provide output information.

Figure 2 shows a summary of the various ways of using the Sagnac Effect to make a rotation sensor. To date, by far the most successful technique is the ring laser gyro, although the schemes whether passive or active, resonant or single-pass, all have very similar inherent sensitivity limits.

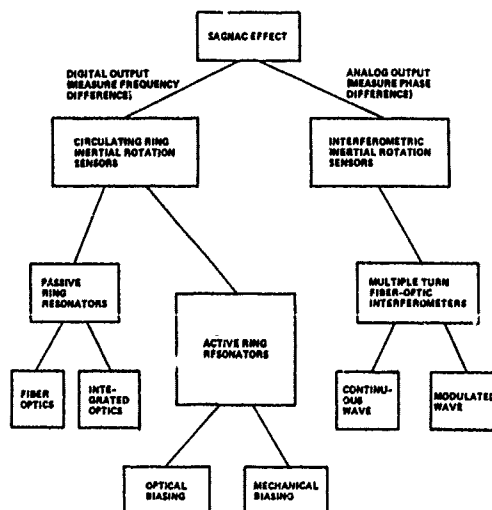
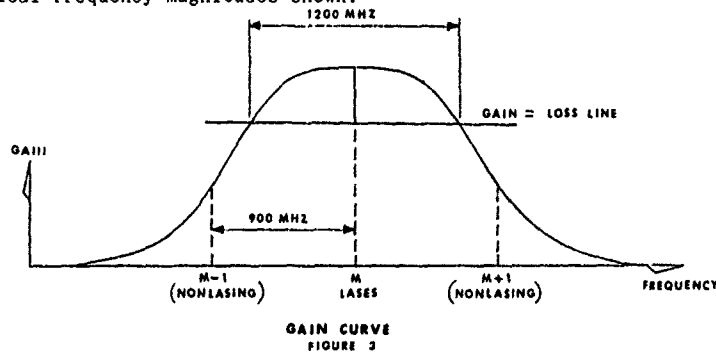


FIGURE 2

THE RING LASER GYRO

Choice of the RLG Gain Medium

For proper operation a ring laser gyro requires a gain medium that sustains monochromatic laser oscillation in two counterpropagating directions simultaneously and also provides maximum opportunity for the beams to interact with the outside world through such effects as backscatter which create gyro lockin discussed later and temperature changes. Primarily for the latter reasons a low pressure gas is the preferred (and currently only) choice. A solid gain medium would necessarily have material interfaces inducing scatter and the high density would invite interaction through temperature and other environmental changes. As mentioned earlier a cavity can resonate at many different frequencies each separated from the next by the cavity free spectral range. Which of these frequencies lase in the cavity is determined by the so-called gain profile of the active medium within the cavity. The low pressure gas provides gain at a range of frequencies centered about the atomic transition responsible for the lasing action. The gain for lasing action is provided by stimulated emission between two energy levels within the medium and in general this transition provides photons which are highly monochromatic. However in the gaseous gain medium of the RLG the atoms or molecules are in constant motion and have a velocity distribution given by a Gaussian profile. Thus the actual frequency of an emitted photon is Doppler-shifted in the frame of the gyro body and depends on the atomic velocity. In this manner the gaseous medium provides gain over a range of frequencies determined by the atomic velocity distribution. This range is temperature dependent but typically is about 1GHz, which is comparable to the free spectral range of a hand-palm-sized ring cavity. The actual range over which lasing can occur is less because only frequencies which have gain greater than the cavity loss can lase. Thus by slight adjustments to the cavity length only one resonant frequency will occur under the gain profile at one time and hence only one cavity longitudinal mode will lase. This situation is illustrated in Figure 3 with some typical frequency magnitudes shown.



GAIN CURVE
FIGURE 3

All ring laser gyros presently use atomic neon gas as a gain medium because the transition is well studied (being the first gaseous lasing medium) and is stable in operation. The gain provided by pure neon is very small and helium gas is added to boost the population inversion via direct helium-neon atom energy transfer due to a useful level coincidence. However even this mixture with naturally occurring neon (about 91% neon 20 and 9% neon 22) will not produce stable oscillation in both directions at peak-gain frequency in a ring cavity. When neon of substantially a single isotope is used, light beams in both directions compete for the same excited atoms with close to zero velocity when the cavity resonant frequency is tuned to the peak gain. Any slight differential cavity loss between the beams will suppress one direction over the other and beam intensities will be very unstable. This problem is overcome by using an approximately equal mix of neon 20 and neon 22. As shown in Figure 4, the peak-gain frequency now occurs between the gain centers for the individual isotopes, and at this frequency the light beams must use a Doppler shift to match the difference in lasing frequency and that of the atomic transition. Since the beams are travelling in opposite directions, they draw on gain atoms with velocities of opposite sign and the mode competition effects are relieved. In practice this scheme produces very stable operation of the counterpropagating beams needed as the basis for all RLG's.

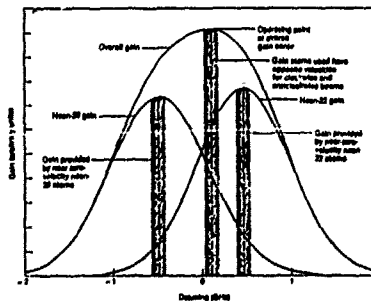


FIGURE 4

Early RLG workers used the 1.15 micron infrared transition in neon because it provided more gain than the more familiar red line. As mirrors became less lossy and cavity alignment methods improved, manufacturers switched to the red line because it is visible and provides a larger gyro scale factor. With current technology, satisfactory lasing action in the red can be obtained with plasma gain lengths of the order of a centimeter or two.

Frequency Locking

The first ring laser gyroscopes were built at Sperry Corporation in the USA in the early 1960's and displayed an unforeseen problem, that of frequency locking or 'lockin'. Any mechanism that couples energy between the counterpropagating beams in the cavity will cause the beat frequency between the beams to be reduced resulting in a nonlinear scale factor between input rate and output beat. More seriously, if the input rate is low enough the two frequencies lock together providing no output beat. A typical plot of the input versus the output in this case is shown in Figure 5. The range of inputs which gives no output is known as the gyro deadband or lockband. The main coupling mechanism in the RLG is the backscattering produced by surface irregularities such as roughness and dielectric changes on the surface of the mirrors within the cavity. Very small amounts of backscatter cause large deadbands; for instance a few parts in 10^{12} of energy from one beam scattered into the direction of the other can create a deadband of several hundred degrees per hour in a palm-sized RLG. A major part of RLG research in the last twenty years has been directed towards methods of circumventing the locking problem. The main approaches are discussed below.

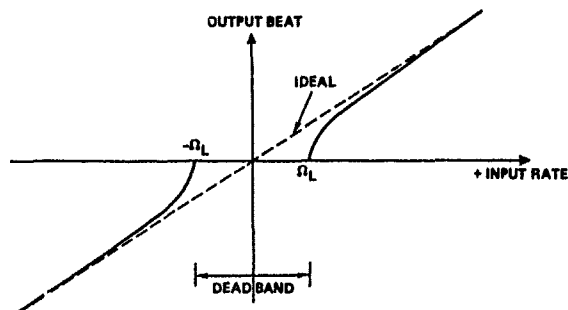


FIGURE 5

There are two popular types of lockin avoidance schemes, those based on so-called d.c. biasing and those based on a.c. biasing. The d.c. schemes rely on providing a fixed bias between the counterpropagating beams of a magnitude large enough to push the deadband well away from the desired operating range of the RLG. Early attempts achieved this with the use of an intracavity glass element with a large Faraday rotation (i.e., Verdet) constant which would thus produce a direction-dependent phase shift for light. The operation of such a device is shown in Figure 6. A glass element which is typically doped with cerium or terbium to increase the Faraday Effect, is placed in an axial magnetic field of several thousand oersted. Circularly polarized light impinging on the glass in the direction of the field will receive an additional phase shift proportional to the field magnitude and glass thickness. Light circularly polarized in the same sense but travelling against the field receives an opposite phase shift, that is, the phase shifts experienced by the beams are nonreciprocal. When such a device is placed in the ring laser cavity the differential phase shift is transformed into a bias between the counter-propagating beam frequencies which may be several MHz.

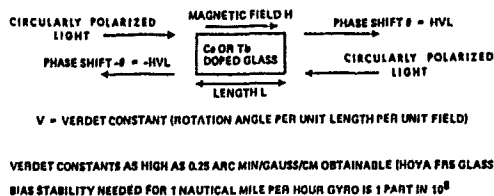


FIGURE 6

These early attempts ran into serious problems with bias stability and were quickly dropped since it appeared that the required magnetic field stabilities of less than a part per million were not achievable. Other d.c. biasing schemes either produced bias shifts that were too small or had similar stability problems.

Interest in d.c. bias techniques was renewed in the late 1960's and early 1970's with the invention of the so-called four-mode gyro or multioscillator. This is an ingenious configuration which allows the operation of two sets of counterpropagating beams in one cavity. The effects of changes in an applied d.c. bias are substantially reduced by common mode rejection when the Sagnac frequency differences for the two beams sets are summed. Recently there has been growing interest in the multioscillator configuration and this scheme is described more fully in a later section.

The a.c. or 'dither' method for overcoming lockin applies a continuously changing bias to the RLG so that the time averaged bias is zero. When the shortcomings of the d.c. scheme became known RLG manufacturers switched to the dithered scheme as their main approach. While this technique can be achieved by the same type of magneto-optical means used in d.c. biasing, the favored approach is to apply a mechanical oscillation to the gyro frame. This has the advantage of being mechanically bounded in average drift. The applied dither continuously sweeps the gyro through the deadband and prevents total lock-up. Analysis using frequency modulation theory shows that, in effect, sinusoidal dither breaks the deadband into a series of much smaller bands occurring at the harmonics of the applied dither frequency. The more energy applied to the dither, then the smaller are the fragments of the deadband. The form of the gyro output curve in the presence of dither is shown in Figure 7 where ω_D is the applied dither angular frequency. The dither is usually applied by means of piezo-electric transducers and is typically of a few arc minutes in amplitude and at a few hundred Hertz. This type of dither approach does add a mechanical component to an inherently mechanical system but avoids the use of intracavity components found in most other schemes, a.c. or d.c. This is probably the main reason for its success; RLG's are so sensitive to scatter effects within the cavity that clear-path devices are seen as being greatly preferred.

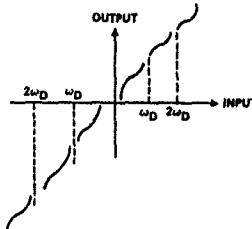


FIGURE 7

Much research has gone into dithered systems and additional electronic circuitry and embellishments have reduced residual error effects. Some efforts have been directed towards non-mechanical a.c. schemes such as the use of Kerr-Effect mirrors that produce a nonreciprocal phase shift on reflection. A magnetic field applied to the back of such a mirror may be continually reversed in sign, driving the Kerr Effect to saturation in each direction thus producing a magneto-optical square-wave dither. Such RLG's have been built and, although clear-path in design, the performance is generally limited by the inferior surface quality of a Kerr-Effect mirror compared to a standard high quality multilayer dielectric stack used for regular RLG's.

The Standing Wave Picture of the Ring Laser Gyro

The ring laser gyroscope stores information about an initial orientation in inertial space, as does its mechanical counterpart, but the means of storage is not as obvious as the orientation of the axis of a spinning rotor. A more pictorial way of examining the RLG is to consider the counterpropagating beams as setting up a standing wave pattern around the cavity as shown in Figure 8. The pattern of bright and dark fringes shown,

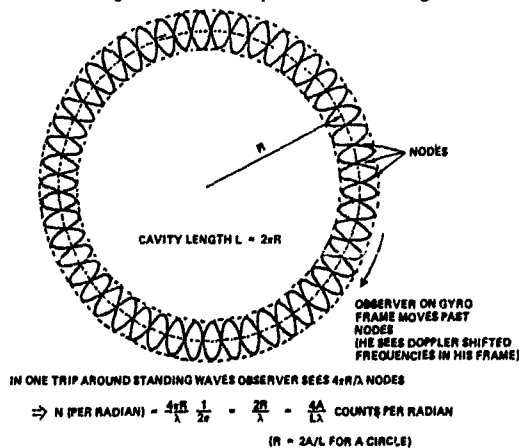


FIGURE 8

corresponding to regions of high and low field energy, actually exist and could be sensed in principle with a detector in the light path. When initially turned on, the light beams set up a wave which is stationary in the inertial frame of the gyro at that instant. The Sagnac Effect governed by the Laws of Relativity forces the wave to remain stationary in this inertial frame and thus allows the pattern to act as a reference for orientation. When the gyro body attached to the host vehicle rotates, the detectors mounted on the mirrors move past the pattern and register the passage of the fringes as gyro counts. Each count corresponds to rotation of the host vehicle by a small angle, typically about one arc second for a palm-sized gyro. In practice of course the light path is not circular as shown in Figure 8 but is triangular or square. In this case the fringes move as though they were interlocked with a stationary fringe pattern inscribing the triangle or square as shown in Figure 9. This visualization will also give quantitatively the scale factor formula of Equation 1 with the fringe spacing equal to half the wavelength of the light in the cavity.

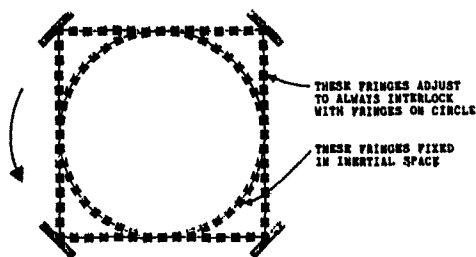


FIGURE 9

The effects of frequency lockin and dither can also be visualized. The mirror backscatter causes the otherwise independent standing wave to interact with the inertial frame of the body and the pattern tends to 'stick' to the mirrors. At total lockin the pattern rotates with the gyro body and no output counts are seen at the detector. When mechanical dither is applied this can be seen as 'shaking' loose the standing wave from the mirrors and allowing the counts to return.

Random Walk in RLG's

One of the error sources of concern in the ring laser gyro is that called random walk. This is a tendency of the device to drift away from its true heading in proportion to the square root of the time advanced and as such is most significant shortly after turn-on. Part of the random walk error is inherent to the gain medium of the laser and is present whatever the configuration. This restriction on performance level is known as the quantum noise limit. Additional random walk may occur as a result of the lockin avoidance scheme, in particular an a.c. biasing scheme.

Firstly consider the effects of the gain medium. As with any gyroscope the RLG holds an initial heading in inertial space as the host vehicle changes its orientation. The RLG stores this information as a phase or frequency difference between the counter-propagating beams and thus any process that scrambles this phase information will create a random walk error. Photons lost from the cavity by absorption or other means are replaced by those from the gain medium. Those photons produced by stimulated emission maintain the phase memory of the initial heading but some are produced by spontaneous emission and add a random component to the gyro output. The ratio of stimulated-to-spontaneous emission is governed by atomic properties of the gain medium but is also proportional to the optical power inside the cavity. Thus the larger the intracavity power, the larger is the fraction of stimulated photons and the smaller is the random walk. Random walk from this source can also be reduced by minimizing cavity losses since this decreases the number of photons required from the gain medium. The random walk of an RLG from gain effects is given approximately by:

$$\sigma/\sqrt{t} = 1.67 \frac{\lambda c}{8\pi A} \sqrt{\frac{h\nu \cdot \text{loss}}{(\text{Power (Internal)})}} \text{ rads}/\sqrt{\text{sec}} \quad (6)$$

where $h\nu$ is the photon energy

Because of the dependence on gyro scale factor the random walk quantum noise limit can always be reduced by making the gyro larger.

The dithered RLG may have an additional source of random walk occurring at the turn-around points in the dither cycle when the gyro is momentarily locked up. This source is proportional to the size of the deadband and uncorrected can be much larger than the random walk from spontaneous emission. So-called dither-turn-around electronics can minimize this error by applying a correction to the gyro output based on the relative beam phases within the cavity at each stationary point within the dither cycle. Other techniques involve manipulating the dither motion to reduce the error.

A typical palm-sized RLG of around 20cm in cavity length may have a random walk of about .0005 to .001 deg/hour from spontaneous emission. Uncorrected dither-noise-induced random walk may increase this by two orders of magnitude. The dither-turn-around correction circuitry or other techniques can restore the random walk to a level approaching that of the quantum noise limit.

Bias Drift Errors

Ring laser gyros are subject to long-term drift errors which in contrast to random walk, build up linearly with time. Apart from the Sagnac Effect there are other physical effects which create nonreciprocal frequency shifts between the counterpropagating beams within the cavity. Most of these produce bias changes that are temperature driven in some way. Some of these are examined briefly below.

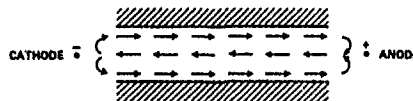
There is a phenomenon known as the Fresnel-Fizeau Effect that will create a frequency bias between the cavity beams if the medium through which they travel is in circular motion with respect to the gyro frame. The Fresnel-Fizeau Effect is actually a consequence of the Doppler frequency shift produced by relative motion and causes the light velocities of the counterpropagating beams to be different to an observer on the frame if the gain medium is flowing in a preferred direction around the light loop. In practice such a flow can be created by the electrical interaction in a gas-discharge between the electrons and ions within the plasma. So-called Langmuir d.c. flow, as shown in Figure 10, creates one such movement although the flow effects in a typical RLG are generally more complex. These flow effects can easily produce biases in the gyro equivalent to many degrees per hour in rotation and are temperature dependent. The basic remedy used by gyro manufacturers to limit these effects is to employ a balanced electrode design with, say, two anodes and one cathode symmetrically placed. In this manner flow effects in one part of the cavity are substantially balanced by opposing flows in another section. For some high accuracy devices aperture and bore layout must be carefully considered as well to achieve the required temperature stabilities.

Another cause of bias drifts in the RLG is from changing magnetic fields. Most RLG's built operate with substantially linearly polarized light beams within the cavity. This is a consequence of a planar light path (whether triangular or square) and the reflection properties of the multilayer-dielectric-stack mirrors used. Linearly polarized light is relatively immune to direction-dependent phase shifts when propagating

VELOCITY OF LIGHT TRAVELING THROUGH A MOVING MEDIUM IS A FUNCTION OF THE VELOCITY OF THE MEDIUM

$$V_L = \frac{c}{n} \pm V_{MED} \left(1 - \frac{1}{n^2}\right)$$

IN GAS LASER MEDIUM IS MOVING BECAUSE OF LANGMUIR FLOW



TO MINIMIZE SUCH EFFECTS USE TWIN ANODE CONFIGURATION

FIGURE 10

through a medium with an applied magnetic field (such as the gain medium in the presence of the Earth's magnetic field). However interactions with cavity mirrors may produce small mode polarization impurities which show up as troublesome nonreciprocal biases even in magnetic fields of less than that produced by the Earth. Thus most RLG's require some form of magnetic shielding.

A ring laser cavity may support lasing modes of a variety of sometimes complex cross sectional intensity distributions across the beam. These are known as transverse modes in contrast to the longitudinal modes separated by the cavity free spectral range. Each transverse mode may lase at a slightly different frequency because of different phase shifts experienced within the cavity. Interactions between these modes may occur and create bias effects, so in general all but one transverse mode are prevented from lasing. The simplest transverse mode called the M_{00} mode, or on-axis mode, has a circularly symmetrical Gaussian intensity distribution. This mode in general has the lowest losses in the cavity if the mirror surfaces are uniform, and the placement of an aperture in the cavity can ensure that only this mode will have low enough losses to lase. For planar cavities this aperture may be elliptical to compensate for the astigmatism introduced by curved cavity mirrors used for alignment stability.

The placement of an aperture in the cavity may itself cause a bias which is sensitive to temperature. This happens when apertures are placed asymmetrically with respect to the gain medium and cause differential gain effects between the counterpropagating beams. Differential scatter effects may also occur in this case, so care is taken in the placement and the shape of the aperture used.

Scale Factor Effects

One advantage of ring laser gyroscopes over their mechanical counterparts is their potential for very linear scale factors over a very large dynamic range. These linearities are measured in parts per billion for the better RLG's. However there are sources of scale factor nonlinearity. The main cause for such deviations is backscatter in the cavity which creates the changes shown in Figure 5 as the gyro mode frequencies are pulled together. This is most evident in the dithered schemes when the gyro continually sweeps through the lockband. In dithered gyros an anomaly occurs in the scale factor and bias at the input rate, called the dither rate, that is equal to the rotation rate at the maximum velocity in the dither cycle. Figure 11 shows the scale factor nonlinearity in a typical gyro for input rates from zero to beyond the dither rate. Clearly it is preferable to make the dither rate as large as possible so that the gyro normally operates below this point where the scale factor error is minimal. This is achieved by putting as

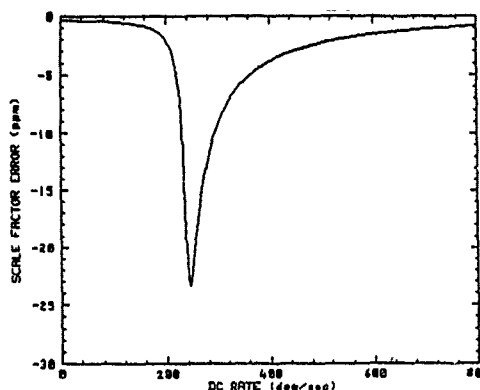


FIGURE 11

much energy as possible into the dither within the mechanical design restraints. Dither rates may exceed 1000 deg/hour. Gyros biased by d.c. means are relatively free from backscatter-induced scale factor effects as long as the bias is large enough. In practice this may require biases of several MHz for nonlinearities less than one ppm.

Scale factor anomalies and shifts can also result from changes in the gain medium, which makes a significant contribution to the optical length of the cavity and as such appears directly in the scale factor of the gyro given by Equation 4. Changes in the refractive index of the gain medium via variations in temperature, plasma excitation, cavity losses and other parameters indirectly create scale factor effects by so-called mode pulling and pushing where the effective resonant cavity frequencies are altered nonsymmetrically. For this reason it is desirable to maintain the lasing mode frequencies at fixed positions near the peak of the gain curve and this entails the use of a servo system that holds the cavity length to within a fraction of a wavelength of light. Such systems are relatively easy to achieve by servoing to the maximum laser output intensity.

The actual enclosed area of the light path may change if tilts in the mirrors or distortions of the gyro frame occur. All such effects can be kept at levels of less than about one part per million with reasonable care in design of the cavity geometry and of the transducers needed to control the cavity length.

Basic Gyro Construction

The modern ring laser gyro is made from a monolithic block of very low expansion glass ceramic such as Zerodur or Cervit, names trademarked by the Schott Glass Company of Germany and Owens Illinois, respectively. The low expansion is desirable in order to minimize changes in the cavity length with temperature. Bore holes and apertures are drilled in the frame to create a light path. The apertures serve to prevent off-axis light modes of more complex intensity cross section than the on-axis Gaussian form required, from lasing. High quality surfaces are polished with critical position tolerances to allow for optical contacting of the mirror.

Mechanically dithered gyros have either a triangular or square light path depending on the preferences of the manufacturer. (As described later the multioscillator gyroscope is unique in requiring a nonplanar light path.) The gyro mirrors are substrates generally made of the same low expansion material as the frame and polished to a very high grade of surface finish to minimize scattering centers from surface roughness. A high quality durable multilayer dielectric stack is coated on the surface. The frame is meticulously cleaned and cathodes and anodes attached for creating the electrical discharge. A means is provided for a fill stem which can be pinched off when the gyro is filled with a suitable gas. The mirrors are attached under stringent clean room conditions to prevent any dust particulates from creating scatter within the cavity. One or

two of the mirrors have slight concave surfaces so that the optical resonator is stable and so that small machining errors in the frame may be corrected by moving these mirrors around for optimal alignment before final attachment.

One or two of the mirrors on the frame are also machined with a thin web structure as shown in Figure 12. When a piezo-electric transducer is attached to the back of the mirror an applied voltage will cause the web to flex and create small changes of an order of the wavelength of light to occur in the path length. This gives enough control to position the cavity resonant frequency in the center of the range of frequencies that receive gain from the plasma discharge. The piezo-electric element are generally made in the form of bimorphs so that the applied voltage causes bending in a manner analogous to a bimetal strip. The transducers are often called pzt's in reference to plumbic zirconium titanate from which they are often made.

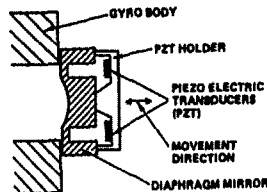


FIGURE 12

The cavity is evacuated to pressures well below one millionth of a torr and baked under these conditions to remove any gaseous impurities absorbed in the walls of the frame. A gaseous mixture of helium and neon is placed in the cavity. The neon may be composite of nearly equal pressures of 20 and 22 isotopes to avoid competition problems as described earlier. The lasing action of the red line in neon is very sensitive to gaseous impurities and so high purity gases are used and much frame processing occurs before the final gas mix is pinched off.

When a high voltage, typically of the order of 1000 volts is applied between the cathode and two anodes counterpropagating beams will lase in the cavity. The beam intensities are monitored outside the cavity via partially transmitting mirrors and a servo control circuit adjusts the cavity length via the mirror piezo-electric transducers to maximize the output intensity. A glass prism known as a combining optic is arranged so that the exiting counterpropagating beams are combined to form a fringe pattern on a pair of semiconductor detectors as is shown in the basic gyro layout of Figure 13. This provides a means for measuring the gyro beat. A pair of detectors is used so that, by arranging them to see the fringe pattern at points spatially separated by one quarter of a fringe, the direction of movement of the pattern and hence of the rotation can be determined.

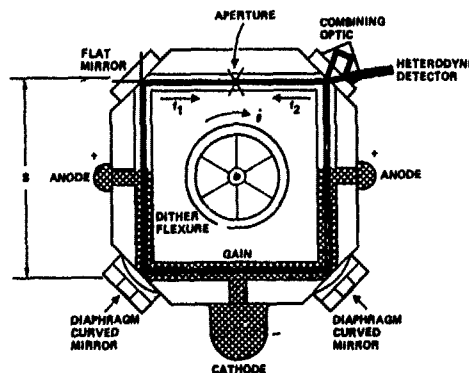


FIGURE 13

The dither mechanism is usually a flexure consisting of several metal vanes spread radially as spokes, each with piezo-electric transducers attached, mounted in the center of the frame as is depicted in Figure 13. An oscillating voltage applied to the transducers induces a vibration in the frame about the central axis relative to the case and host vehicle. Electronic circuitry working in conjunction with a position pick-off that registers the position of the gyro, extracts the applied dither from the gyro output leaving the required rotation rate of the host vehicle. Some configurations use a beat detector mounted to the gyro case rather than the dithering frame to remove the dither component.

The gyro is usually hermetically sealed with some control electronics in a case to prevent problems with moisture. The case material is often of a high μ material to provide some amount of magnetic shielding. A three-axis system is formed when three such gyros are mounted orthogonally in a triad along with the necessary electronics for running each gyro. Some designs incorporate all three axes in one block of glass; for instance, by using a cube with a mirror attached to the center of each face, three orthogonal light paths can be obtained with only six mirrors.

The Multioscillator RLG

The success of the mechanical dither scheme for overcoming lockin brought the first commercial RLG systems onto the market in the late 1970's and most manufacturers devote a major part of their effort towards improving these systems. However the mechanically dithered gyro is seen as having several drawbacks. The dither itself acts as a source of mechanical noise and vibration on the host vehicle. Additional electronics is needed to create the dither and minimize residual lockin effects, and the full scale factor linearity and random walk performance of the RLG is limited by this form of lockin circumvention.

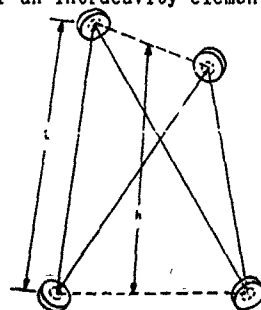
An alternative mechanical means is to 'carousel' the gyro; that is to rotate it on a turntable for several revolutions before reversing the direction. This is a mechanical d.c. biasing scheme and involves additional component complexity but is being tried for military systems where the mechanical dither cannot be tolerated.

Magneto-optical biasing schemes are probably most in keeping with the optical nature of the RLG. As discovered early in RLG development such schemes generally have several serious setbacks. The problem with stringent conditions on magnetic field stability killed early interest in such ideas. In addition use of the magneto-optical effect usually requires some form of intracavity solid element, which enhances scatter effects, and dictates the use of circularly polarized light in the cavity which may be difficult to produce and is inherently more sensitive to external magnetic fields.

Apart from some efforts with a.c. schemes using magnetic Kerr-Effect mirrors mentioned earlier, magneto-optical biasing in RLG's has centered around the use of a four-mode scheme known as the multioscillator approach. This was derived in the late 1960's to obviate the need for a stable magnetic field bias. The basic idea is to operate two independent counterpropagating beam pairs in the same cavity so that they each receive a d.c. bias from the same element of exactly equal magnitude but of opposite sign. Then by adding the output beats from the two gyros the Sagnac contribution is doubled while the bias contribution cancels.

The multioscillator RLG is possible because a ring optical cavity will in general resonate with light of two different polarizations in each direction simultaneously. In a cavity with no losses these light polarizations are orthogonal. The resonant frequencies of each polarization are not necessarily equal and depend on the cavity geometry plus the polarizing characteristics on reflection of the mirrors making up the cavity. Most multilayer-dielectric-stack mirrors behave similarly to a single dielectric interface on reflection; namely linearly polarized light in the plane of incidence (p-type light) and that polarized normal to the plane of incidence (s-type light) receive a differential phase shift of 180 degrees. With these types of mirrors a triangular cavity will lase with p-type modes shifted in light frequency by one half a free spectral range (or about 500 MHz in a 30cm cavity) from the s-type modes; in practice the cavity is tuned in length so that only the lower-loss s-type light lases. A square cavity with these mirrors would lase with s- and p-type linear modes at the same frequency but mode competition effects combined with the lower losses of the s-type mode suppress the p-type mode altogether. Thus these types of planar-geometry gyro operate only with s-type linearly polarized light and are called two-mode gyros.

The multioscillator cavity is arranged so that left and right circularly polarized modes will resonate at different frequencies with substantially the same cavity losses. The frequency difference between the two polarizations is known as the reciprocal splitting. Both the circularly polarized modes and the reciprocal splitting can be achieved by using an intracavity element such as crystalline quartz with its optical axis aligned with the light propagation direction. However modern multioscillator cavities use an out-of-plane geometry shown schematically in Figure 14 which accomplishes the same goal but without the need for an intracavity element. This realization was a major step in



L = CAVITY LEG LENGTH
h = CAVITY HEIGHT
S = ASPECT RATIO
 $\frac{L}{h} = S$

FIGURE 14

four-mode gyro technology. The reciprocal splitting is a function of the degree of nonplanarity; a plot in Figure 15 shows how the effective reciprocal splitting between closest modes varies with the aspect ratio (height over leg length) for a four-mirrored equal-leg-length cavity.

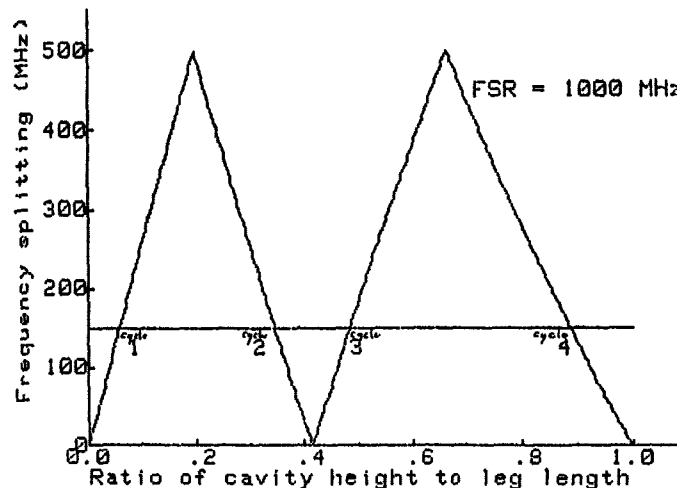


FIGURE 15

Each polarization that lases in the multioscillator cavity has two directions of propagation and so creates an independent gyro pair. To prevent lockin within each pair a nonreciprocal bias must be applied and this generally takes the form of a thin glass material with relatively large Verdet constant. When an axial magnetic field is applied to the element the counterpropagating beams are split in frequency as described earlier and shown in Figure 6. The important fact that makes the multioscillator practical is that the left circularly polarized modes are split in the opposite direction to the right circularly polarized modes thus giving the mode order shown in Figure 16(A). The output beat of the multioscillator is taken as the sum of the individual gyro beats and is:

$$(f_{La} - f_{Lc}) + (f_{Ra} - f_{Rc}) \quad (7)$$

Magnetic field variations that change the nonreciprocal bias are cancelled out as shown in Figure 16(B) by:

$$\begin{aligned} & [(f_{La} + \Delta f_H) - (f_{Lc} - \Delta f_H)] + [(f_{Ra} + \Delta f_H) - (f_{Rc} - \Delta f_H)] \\ & = (f_{La} - f_{Lc}) + (f_{Ra} - f_{Rc}) \end{aligned} \quad (8)$$

However the Sagnac Effect produces twice the beat change compared to the two-mode gyro as shown in Figure 16(C) by:

$$\begin{aligned} & [(f_{La} - \Delta f_r) - (f_{Lc} + \Delta f_r)] + [(f_{Ra} - \Delta f_r) - (f_{Rc} + \Delta f_r)] \\ & = (f_{La} - f_{Lc}) - (f_{Ra} - f_{Rc}) - 4\Delta f_r \end{aligned} \quad (9)$$

The multioscillator gyro, particularly in its nonplanar-cavity configuration, is an interesting device in that it not only solves the problem of d.c. bias stability but also provides a device with twice the scale factor of an equivalently sized two-mode gyro. The random walk is also improved by a factor of $\sqrt{2}$ over a two-mode gyro of the same size.

The main drawback of the multioscillator as described is that it requires an intra-cavity glass element that not only creates cavity losses and scatter, and complicates manufacture but more recently may limit the device's performance in a nuclear environment since nuclear radiation tends to darken the glass. Despite these disadvantages, recent developmental efforts with the multioscillator indicate that it is a viable inertial rotation sensor.

Early multioscillator workers in the 1970's attempted to remove the glass intracavity element and use the gain medium itself as a means of providing a nonreciprocal bias. Normally gaseous media will give only very small Faraday rotations and would require impractically high magnetic fields to give the required splitting of at least several hundred kilohertz. However the neon gas in the gain medium is acting at atomic resonance

and is capable of producing splittings of this size with axial magnetic fields of a few hundred Oersted (i.e., fluxes of a few hundred Gauss). Multioscillators that use a magnetic field on the plasma rather than a separate glass element became known as Zeeman laser gyros or ZEELAG's.

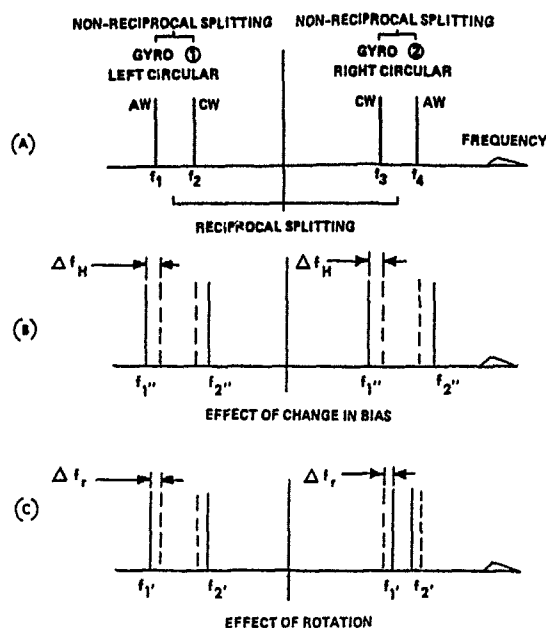


FIGURE 16

The ZEELAG never lived up to its initial promise and displayed large sensitivities to changes in cavity length and temperature. These arise from the fact that the Verdet constant and hence the amount of nonreciprocal splitting produced by the gain medium is a strong function of light frequency. As shown in Figure 17, the left and right circularly polarized gyros operate on opposite sides of the curve and thus experience opposite changes in magnitude when the cavity length is changed, giving rise to large detuning sensitivities. With this realization multioscillator manufacturers switched back to the intracavity glass element.

In the future it seems possible that laser gyro designers will find other forms of the multioscillator that will fill the void for a non-dithered ring laser gyro. Currently multioscillator devices with thin intracavity glass elements in development are producing results that indicate commercial inertial sensors are not far away.

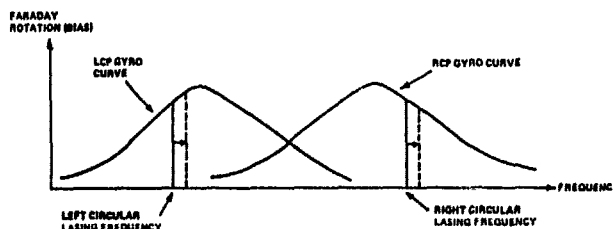


FIGURE 17

INERTIAL GRADE FIBER GYROS

by

G.A. Pavlath
 Litton Guidance & Control Systems
 5500 Canoga Avenue
 Woodland Hills, CA 91367-6698
 United States

SUMMARY

Fiber gyros are being developed for numerous applications by many people around the world. These applications include smart munitions, tactical missiles, attitude and heading reference systems (AHRS), inertial navigation, and tracking and pointing. This paper will review existing methods which have been developed for achieving inertial navigation grade performance in fiber gyros. Recent data obtained at Litton on an engineering development model of an inertial grade fiber gyro will also be presented.

I. INTRODUCTION

One of the more remarkable developments in modern technology is the displacement of spinning wheel gyros by instruments which measure angular rotation using only optical means. The ring laser gyro, after nearly twenty years of intense development, has emerged as a superior candidate for many inertial applications.

Presently, the fiber-optic gyro has attracted widespread attention as a potential alternative to ring laser gyros. Its apparent simplicity - a laser diode, a photodetector and a coil of inexpensive fiber - has led many to believe that it offers an easy way to avoid the investment in high technology development which in the past has been a prerequisite to successful competition in inertial navigation.

Despite its apparent simplicity, the fiber-optic gyro is a highly sophisticated instrument with many subtle error sources which must be understood and controlled. Fortunately, a respectable period of advanced development has already occurred and we have arrived at a state of the art where inertial navigation performance instruments have been demonstrated in a laboratory environment. Inertial navigation grade fiber gyros will be available in the near future.

II. HISTORY

The Sagnac effect was discovered by the French scientist, G. Sagnac, in 1913.¹ Michelson and Gale used the Sagnac effect to measure the rotation of the earth in 1925.² Their apparatus used evacuated sewer pipes to form a closed optical path with a perimeter of 1.2 miles. The Sagnac effect remained a laboratory curiosity for many decades after this demonstration because of its low rotation sensitivity.

The fiber-optic gyroscope was first proposed by Vali and Shorthill³ in 1976. They predicted shot noise limited sensitivity of 10^{-3} deg/hr but were able to demonstrate only about 2 deg/sec. In 1977, Arditty, Shaw, and Chodorow of Stanford University demonstrated⁴ a rotation sensitivity of 0.4 deg/sec. Davis and Ezekiel⁵ demonstrated, in 1978, a technique for increasing the sensitivity of fiber-optic gyros and they obtained a rotation sensitivity of about 130 deg/hr. The first practical fiber-optic gyro was demonstrated by Cahill and Udd⁶ in 1979. Using a closed loop mechanization, they obtained a rotation sensitivity of 0.5 deg/sec and a greatly increased dynamic range.

In 1980, Ulrich identified the importance of reciprocity in fiber-optic gyros.⁷ By properly arranging the components in a reciprocal configuration, he was able to obtain a rotation sensitivity of 3 deg/hr. Also in 1980, coherent Rayleigh backscattering was identified as the major limitation of fiber gyro sensitivity by Cutler, Newton, and Shaw.⁸ All of the previous fiber gyros had used a combination of bulk optic components with a fiber-optic sensing coil. Bergh, Lefevre, and Shaw built the first all fiber gyroscope in 1980.⁹ Their gyro was fabricated entirely from one single unspliced fiber. By 1981, they were able to demonstrate rotation sensitivities of 0.2 deg/hr.¹⁰ Davis and Ezekiel reported closed loop operation of a fiber-optic gyro with rotation sensitivity of 0.1 deg/hr in 1981.¹¹ Ezekiel, Davis and Hellwarth identified the Kerr effect as a major source of bias uncertainty shortly thereafter.¹² By 1982, the Kerr effect had been eliminated as a source of bias uncertainty by Bergh, Culshaw, and Shaw.¹³ The bias uncertainty on their all-fiber gyro was below 0.04 deg/hr and the angular random walk was below 0.001 deg/root-hr.¹⁴ In 1983, Burns, Moeller, Villararuel, and Abbebe of the Naval Research Lab. demonstrated a bias uncertainty of 0.01 deg/hr with a simplified mechanization using special birefringent fiber.¹⁵

In 1984, Arditty, Lefevre, and Graindorge reported on a closed loop fiber gyro which achieved a three sigma bias error of 0.1 deg/hr and a scale factor linearity of 100 ppm.^{16,17,18} This gyro achieved closed loop operation without the use of a frequency shifter. The gyro used an integrated optics phase modulator driven with a digital ramp voltage. This technique has been referred to as the digital serrodyne or as the quantized phase ramp.

III. PRINCIPLES OF OPERATION

Optical gyroscopes, ring-laser and fiber-optic gyros all utilize the Sagnac effect to measure rotation. The Sagnac effect is a relativistic rotation-induced time difference between optical waves which counterpropagate around a closed optical path. Ring-laser and fiber-optic gyros are two different ways of mechanizing the measurement of this rotation-induced time difference, ΔT , which is

$$\Delta T = \frac{4LR}{c^2} \Omega \quad (1)$$

where L is the length of the path, R is the radius of the path, Ω is the angular velocity and c is the speed of light in free space. Equation (1) is correct to first order in $R\Omega/c$ and, for all physically realizable rotation rates, is accurate to about 1 part in 10^{17} . For an in-depth treatment of the Sagnac effect, see references 19-21.

Two principle mechanisms exist for measuring the Sagnac effect: resonant and interferometric. Most work to date in fiber gyros has focused on the interferometric mechanization as it is viewed to be nearer production than resonant mechanizations. Recently, several groups have started examining various resonant mechanizations of fiber gyros for longer range applications (22,23). Since this paper is concerned with the nearer term, inertial grade applications of fiber gyros, it will focus only on the interferometric mechanization.

The fiber optic interferometric implementation of the measurement of the Sagnac effect is shown schematically in figure 1. In this mechanization the Sagnac time difference is converted into a phase shift between counterpropagating optical waves. Light from the source is split into two parts by a beam splitter or a fiber-optic directional coupler^{24,25} and each part propagates in opposite directions once around the closed optical path. The waves are then recombined and the intensity of the combined waves is measured with a photodetector. The photocurrent, i , is related to the phase difference, ϕ , between the counterpropagating optical waves as

$$i = i_0[1 - \cos(\phi)]/2 \quad (2)$$

where i_0 is the peak photocurrent. The phase difference between the optical waves is proportional to the time difference which results from the Sagnac effect. This phase difference, which hereafter will be referred to as the Sagnac phase shift (ϕ_s), is

$$\phi_s = 2\pi LD\Omega/\lambda c \quad (3)$$

where L is the length of the closed optical path, D is the diameter of the optical path, Ω is the angular velocity, λ and c are the free space values of the optical wavelength and speed of light respectively. By measuring the photocurrent and applying equations (2) and (3), the angular velocity can be determined.

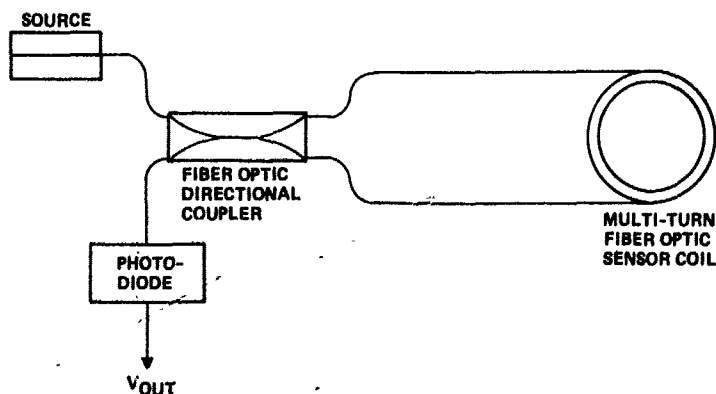


Figure 1 - Basic Sagnac ring interferometer.

The transfer function of the fiber interferometric gyro is plotted in figure 2a. It is seen that the rotation sensitivity at rest is zero and that the direction of rotation cannot be determined. To achieve maximum rotation sensitivity and to resolve the direction of rotation,⁵ the gyro is configured as shown in figure 3. The phase modulator

in the sensing coil sinusoidally modulates the phase difference between the CW and CCW optical waves. The instantaneous photocurrent is

$$i(t) = i_0[1 - \cos(\theta_s + \theta_m \cos \omega t)]/2 \quad (4)$$

where ω is the modulation frequency and θ_m is the amplitude of the phase modulation. Synchronously demodulating the photocurrent in phase with the applied modulation results in the following output voltage

$$v = v_0 J_1(\theta_m) \sin(\theta_s) \quad (5)$$

where J_1 is the first order Bessel function of the first type and v_0 is the peak voltage output. This transfer function, plotted in figure 2b, exhibits maximum rotation sensitivity at rest and the direction of rotation is resolvable. The effect of this modulation is to introduce a 90 degree phase bias between the CW and CCW optical waves. This phase bias is exact and stable and occurs for any modulation frequency. In addition, the zero of the transfer function is now independent of the optical power incident on the photodiode and the electronic gain.

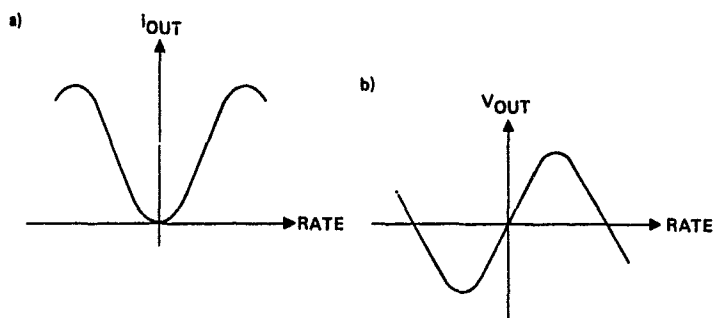


Figure 2 - Transfer function of a Sagnac interferometer, a) basic interferometer, and b) phase-modulated interferometer.

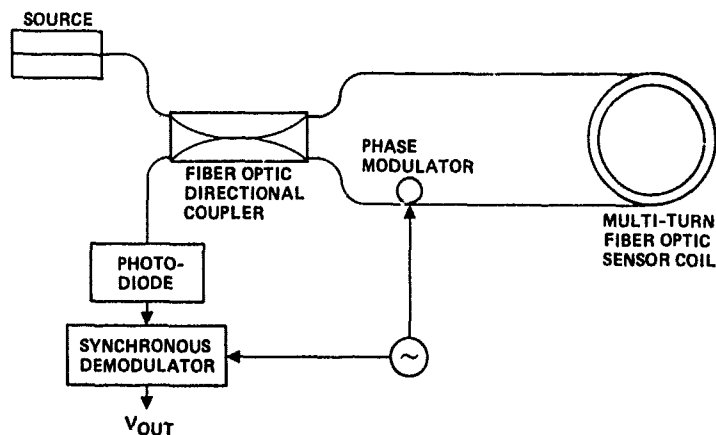


Figure 3. - Schematic of a Sagnac gyro which uses phase modulation and synchronous demodulation to achieve maximum rotation sensitivity at rest.

IV. ELIMINATION OF ERRORS

The development of fiber gyros is characterized by the identification and elimination of a multitude of limiting error sources. Initial efforts were directed at reducing short term rate noise in the gyros to achieve the predicted rotation rate sensitivities.³ When this was accomplished, researchers next examined and eliminated the sources of bias

uncertainty. Presently, development efforts are focused on increasing the linearity of the instrument and improving the scale factor stability.

Sources of Noise

The principal sources of short term rate error noise in fiber-gyros are shot noise and coherent optical backscattering.

The fundamental limitation on rotation sensitivity is due to shot noise which is the quantum mechanical uncertainty in the number of carriers generated in the photodiode. The shot noise is proportional to the square root of the optical power incident on the photodiode. Presently, the magnitude of all other noise sources has been reduced below that of the shot noise, and several fiber-optic gyros have reported shot noise limited angular random walks of less than 0.001 deg/root-hr.^{14,26}

Light propagating in the fiber gyro is scattered at discrete points, such as splices and joints, due to Fresnel scattering, and continuously along the fiber due to Rayleigh scattering. A portion of this scattered light is captured by the fiber and propagated in the backwards direction. Four optical waves, two forward traveling waves which contain the rotation rate information and two backscattered waves whose phase compared to the signal waves is randomly varying, are coherently superposed at the photodiode. This coherent superposition results in a random error in the measurement of the rotation rate.⁸ The magnitude of this error is proportional to the square root of the back-scattered optical power which is coherent with the forward signal waves. Several methods have been proposed to reduce this error source.^{8,9,27} A practical technique is to use an optical source with a short coherence length^{9,14} such as a laser diode or a super luminescent diode (SLD).

Sources of Bias Uncertainty

The principle sources of bias uncertainty in a fiber gyro are polarization nonreciprocity, time varying phenomena (e.g., Schupe effect), Faraday effect, Kerr effect, and phase modulator imperfections.

Reciprocity means that, in the absence of rotation, the CW and CCW optical paths thru the gyro are identical. The total optical phase shift in either the CW or CCW direction is approximately 10^{-10} radians. To achieve a rotation sensitivity of 0.01 deg/hr requires the measurement of the phase difference between the CW and CCW waves of 10^{-7} radians. This means that the CW and CCW optical paths must be identical to better than 1 part in 10^{-17} in the absence of rotation. The fact that this equality is routinely achieved today is truly astounding. The fiber gyroscope schematically depicted in figure 3 is not reciprocal. To insure reciprocity requires the addition of a second directional coupler, a spatial filter, and a polarizer as depicted in figure 4. The details of how the non-reciprocity occurs and how it can be eliminated are beyond the scope of this paper. References 4, 7 and 28 should be consulted for a thorough treatment of this topic.

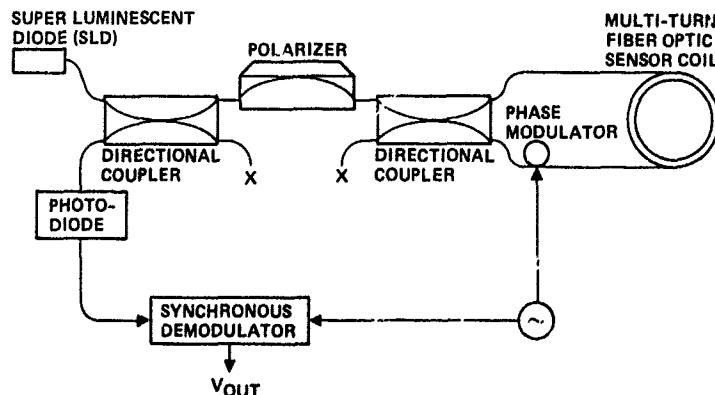


Figure 4 - Schematic of a reciprocal fiber-optic Sagnac gyro.

An ideal polarizer is a filter which passes one state of polarization without loss while completely attenuating the orthogonal state of polarization. A real polarizer does not entirely attenuate the orthogonal state of polarization.

The extinction ratio is defined as ten times the logarithm of the ratio of the power in the orthogonal polarization at the output of the polarizer to that at the input and is a measure of the quality of the polarizer. For the fiber gyro depicted in figure 4 to be

completely reciprocal requires that the extinction ratio be infinite. The finite extinction ratio of realizable polarizers causes a small non-reciprocity which results in a bias. The bias is unstable because the non-reciprocity is environmentally sensitive. The magnitude of the bias uncertainty is proportional to the square root of the extinction ratio and is treated in detail in Reference 29. To reduce the bias uncertainty to 0.01 deg/hr requires an extinction ratio greater than 100dB and some form of active polarization control in ordinary telecommunications fiber.

Fiber gyros which use a special type of fiber (e.g., polarization maintaining or PM fiber), along with a low coherence optical source (e.g., a superluminescent diode) can significantly reduce the extinction ratio requirement of the polarizer for inertial navigation accuracies. Such fiber gyros^{15,26} have demonstrated bias stabilities of 0.01 deg/hr with extinction ratios of only 60 dB.

Reciprocity is strictly valid only for time invariant systems. When time varying phenomena, such as thermal gradients and acoustic waves, are present, additional phase shifts between the CW and CCW waves result in rotation rate measurement errors.³⁰ The effect of these errors can be reduced several orders of magnitude by winding the fiber with a special so-called "quadrupole" technique.³¹ The effect of this winding technique is to render the application of the time varying phenomena symmetric about the midpoint of the fiber. The phase shifts now induced by these phenomena are common mode to the CW and CCW waves to first order and are rejected at the photodiode. For an in-depth treatment of this error source, consult References 14, 30 and 31.

When an optical fiber is exposed to a magnetic field, a phase shift between the CW and CCW optical waves occurs due to the Faraday effect. This phase shift, ϕ_f , is

$$\phi_f = V \oint p H dl \quad (6)$$

where V is the Verdet constant of the fiber, p is a function of the state of polarization, H is the magnetic intensity, and dl is a differential of length along the direction of propagation. The line integral in equation 6 is non-zero, even though there are no lines of current through the fiber-optic sensing coil because p is a function of position along the fiber. This magnetically induced phase shift between CW and CCW waves results in a bias which is unstable due to the environmental sensitivity of the function p . This source of bias instability can be reduced by magnetically shielding the gyro. References 32 and 33 provide a good treatment of the Faraday effect in fiber gyros.

The Kerr effect^{12,13,34,35,36} is a third order optical non-linearity which occurs in optical fibers. Its effect in the fiber gyro is to produce a phase shift, which results in a bias, between the CW and CCW waves proportional to the difference in the intensities of the waves. The difference between the CW and CCW intensities is determined by the coupling ratio of the directional coupler in the fiber-optic sensing coil. It is not practical to stabilize the coupling ratio to the required degree to obtain bias uncertainties of 0.01 deg/hr. It has been shown that the Kerr effect can be suppressed by appropriately modulating the source³⁷ or by using a source, such as a SLD, with the correct statistics.^{13,36}

The phase modulator, which is used to bias the gyro to maximum rate sensitivity at rest, modulates not only the phase of the optical wave but also the polarization to a small degree. The polarizer acts as a discriminator and converts the polarization modulation to an intensity modulation. This intensity modulation is at the detection frequency and results in a bias. The bias is unstable because the magnitude of the polarization modulation critically depends on the state of polarization of the wave at the modulator, which varies randomly. It has been shown,¹⁰ that if the modulation frequency is properly chosen then the bias due to polarization modulation can be eliminated. At this particular modulation frequency, the polarization modulation of the CW and CCW waves is 180 degrees out of phase and the resulting intensity modulation of one wave is cancelled to first order by that of the other wave.

Sources of Scale Factor Error

The output of the fiber gyro, shown in figure 4, is sinusoidal in the input rate (figure 2b). Linear operation occurs only for small rate inputs and hence the dynamic range of the gyro is restricted to about 5 orders of magnitude. The scale factor of the gyro depends on a multitude of parameters (e.g., source optical power, coupling ratios of the directional couplers, state of polarization, etc.). These parameters cannot be controlled sufficiently to achieve the scale factor stability required. The fiber gyro, depicted in figure 4, is essentially a null sensor and is suitable for gimballed platform applications.

Strapdown aircraft inertial navigation requires an instrument dynamic range of 8 to 9 orders of magnitude and a scale factor stability of a few ppm. To meet these requirements, a closed loop phase compensation technique must be implemented in the fiber gyro. Techniques which have been proposed are differential frequency propagation,^{6,11,38,39} phase modulation,^{16,17,18,40,41} and the use of Faraday phase shifter.⁴²

These phase compensation techniques all seek to introduce a controllable, non-reciprocal phase shift, ϕ_{nr} , between the CW and CCW optical waves. The output of the synchronous demodulator becomes:

$$v = v_0 J_1(\phi_m) \sin[\phi_s + \phi_{nr}] \quad (7)$$

The amount of non-reciprocal phase shift required to null the demodulated voltage is the measure of the rotation rate and has the required linear transfer function.

The differential frequency propagation method was the first closed loop technique demonstrated on fiber gyros and achieved scale factor accuracies of about 60 ppm. For the gyro depicted in figure 5, the frequency of the CCW wave is shifted prior to propagation through the fiber-optic sensing coil while that of the CW wave is shifted after propagation through the coil. Propagating through the coil at different frequencies results in a non-reciprocal phase shift proportional to the frequency difference. The rotation rate can be obtained by use of equation (8), where ΔF is now the frequency of the output of the VCO, D is the diameter of the fiber sensing coil, λ is the wavelength of the source, n is the index of refraction of the fiber, and Ω is the angular rate.

$$\Delta F = \frac{D}{\lambda n} \Omega \quad (8)$$

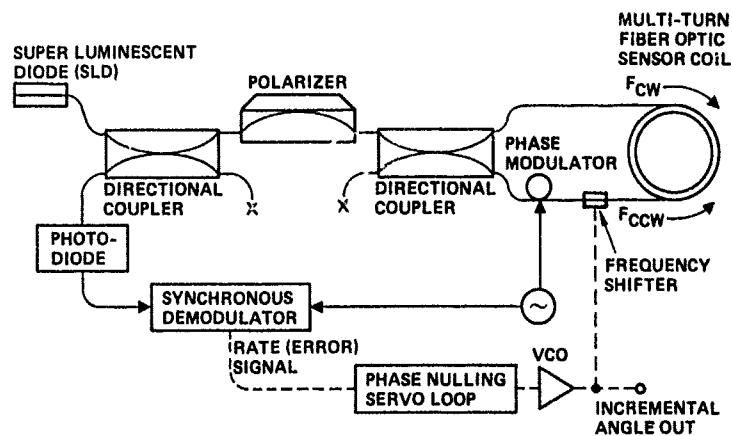


Figure 5 - Schematic of a closed loop, phase-nulling Sagnac fiber-optic gyro which uses differential frequency propagation.

In addition, each zero crossing of the VCO output corresponds to a fixed angular increment of the instrument. The differential frequency propagation technique has converted the fiber gyro from a rate instrument into a rate integrating instrument.

The digital serrodyne or quantized phase ramp is the newest technique¹⁶ for closing the loop on the fiber gyro. Scale factor accuracies of 100 ppm have been achieved^{17,18}. A closed loop fiber gyro utilizing the quantized phase ramp technique is schematically depicted in figure 6. A square wave modulation technique very similar to the sine wave technique discussed previously is used to dynamically bias the gyro for maximum rotation sensitivity.

To this bias signal a digital staircase voltage depicted in figure 7a is added. The duration of each step is matched to the transit time (τ) of the fiber sensing coil. The CCW optical wave sees a phase shift proportional to the modulator constant (K) times the step voltage (V). The CW wave sees the phase shifter τ seconds later. It can be shown that the result of this phase modulation is the nonreciprocal phase shift depicted in figure 7b which adds to the Sagnac phase shift.

For the majority of the time the nonreciprocal phase shift equals the modulator constant times the step voltage. The first servo loop adjusts the step voltage so that the Sagnac phase shift is continuously cancelled except during the period in which the staircase voltage is reset. The frequency of resets is now proportional to the angular rate of the instrument as given by equation 8 and each reset equals a fixed angular increment given by equation 9.

$$\theta = \lambda n / D \quad (9)$$

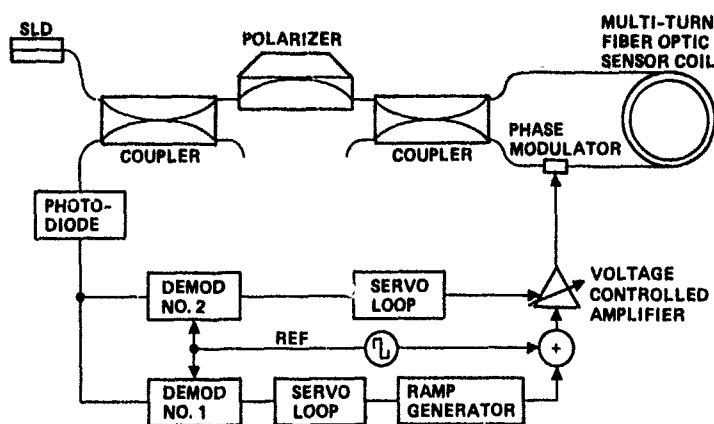


Figure 6 - Schematic of a closed loop fiber gyro which uses the quantized phase ramp technique.

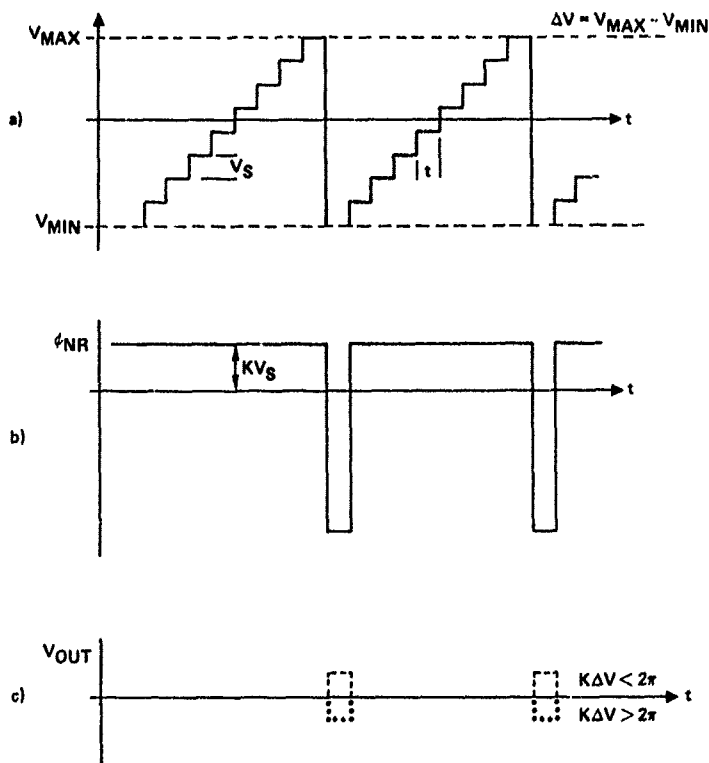


Figure 7 - Phase modulator drive (a), non-reciprocal phase shift (b), and photodiode output voltage (c) in quantized phase ramp fiber gyro of figure 6.

In equation 9, λ is the source wavelength, n is the index of refraction of the fiber, and D is the diameter of the fiber sensing coil.

The scale factor of the fiber gyro is now dependent on the accuracy with which the reset voltage (i.e., difference between V_{max} and V_{min}) is held constant. The key discovery made by researchers at Thomson CSR^{17,18} was that this voltage could be stabilized to great accuracy using the fiber gyro itself. Figure 7c depicts the output voltage of the photodetector when the first servo is closed about the gyro. A signal is present during the resets of the digital phase ramp proportional to the deviation of the reset voltage from a 2π phase shift in the fiber gyro. The second demodulator demodulates this signal and through a servo loop controls the gain of the amplifier, through which the modulation and ramp voltages are applied, such that the reset voltage equals 2π radians of phase shift.

The scale factor of a closed loop fiber gyro still depends on the wavelength of the source. If the required scale factor stability is a few ppm, then the center wavelength of the optical source must be stable to at least 1 ppm. The scale factor non-linearity due to dispersion in the optical fiber is treated in Reference 39.

V. EXPERIMENTAL RESULTS

Litton Guidance and Control Systems has been actively engaged in the development of inertial navigation grade fiber gyros since early 1982. Experimental results obtained in a laboratory environment on a closed loop fiber gyro will now be presented.

Figure 8 shows the fiber gyro on which the experimental results were obtained. The gyro is 4 inches in diameter and 3/4 of an inch high. The source and detector were externally mounted on the electronics cards (not shown) which were fabricated from commercially available discretes.

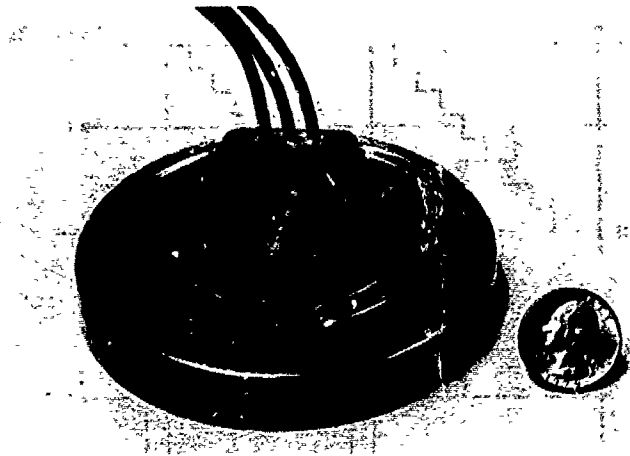


Figure 8 - Litton Inertial Grade Fiber Gyro Developmental Model (EDM-2)

Figure 9 shows typical bias stability performance in a laboratory environment. The lower curve depicts the internal instrument temperature. The temperature swing results from the laboratory air conditioning system. The fiber gyro was not thermally controlled or isolated for any of the measurements. The upper curve represents the raw, uncompensated gyro output achieved under closed loop operation. The standard deviation for this run is 0.013 deg/hr which corresponds to an uncertainty of 0.9 counts. The random walk measured from this drift run is less than 0.005 deg/rt-hr and is measurement limited due to quantization. No fast filter was available to extract the real random walk of the instrument which is estimated to be about 0.001 to 0.002 deg/rt-hr.

Figure 10 shows the day-to-day, turn-on-to-turn-on bias repeatability of this instrument. During the course of this measurement the gyro was repeatedly unhooked from its closed loop electronics, demounted from the rate table and stored on a shelf, re-mounted, reconnected to its electronics without any adjustment of the electronics, and the bias measured. The one sigma uncertainty in the bias is 0.0028 deg/hr with closed loop operation. The data in the figure is again raw, uncompensated data and was achieved without any thermal control of the gyro under test.

The above data is believed to represent the best fiber gyro bias stability and repeatability data obtained to date. It is significant that the data was obtained using closed loop electronics and without any temperature control.

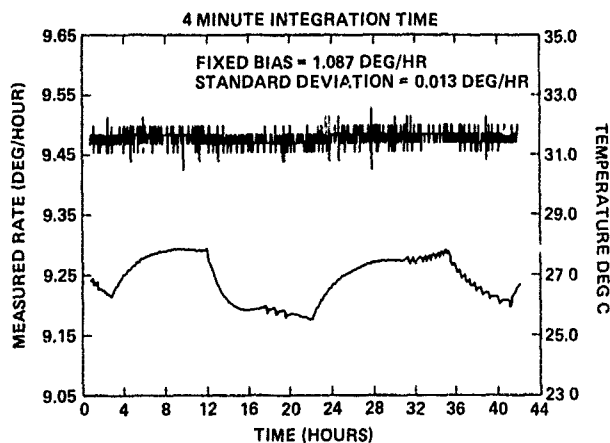


Figure 9 - Typical Bias Stability Measurement (Uncompensated).

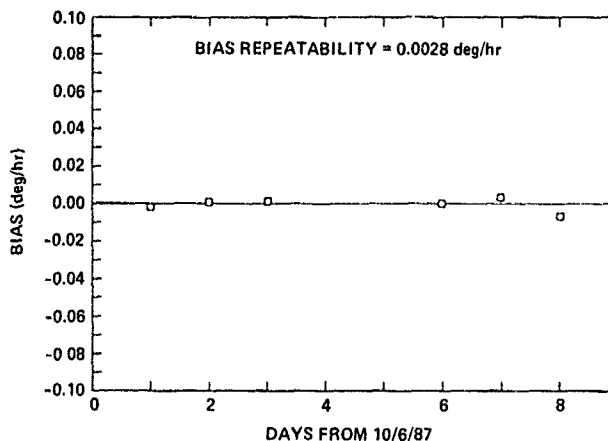


Figure 10 - Day-to-Day, Turn-on-to-turn-on Bias Repeatability (Uncompensated).

During the bias measurements reported above, dynamic testing of the gyro was also conducted to determine scale factor stability and linearity. Figure 11 depicts the raw data obtained during a typical dynamic test. The scale factor derived from this data is about 3.55 arcsec/count and agrees very well with the design value. Figure 12 depicts the deviation of the previous data from the ideal transfer function. Scale factor linearities of less than 10 ppm (one sigma) of full scale have been achieved.

The prime source of scale factor error is the change in optical wavelength of the source. The scale factor of the instrument is linearly dependent on the wavelength of the source. The source wavelength changes strongly with temperature (approximately 200 ppm/deg-C). An experiment to measure this source of scale factor error was performed. Figure 13 depicts the measured wavelength induced scale factor error from -10 to +50 deg-C. The peak scale factor error over this range is 20 ppm and is believed to be instrument limited. A time lag between the temperature measured by the thermal probe and the gyro under test is felt to cause the hysteretic behavior. Actual scale factor error from this source over temperature is felt to be better than 20 ppm.

The above data is believed to represent the best dynamic data reported on a fiber gyro to date. It is important to realize that a single fiber gyro with a single set of electronics without any adjustments whatsoever was simultaneously able to achieve good noise, bias and scale factor performance.

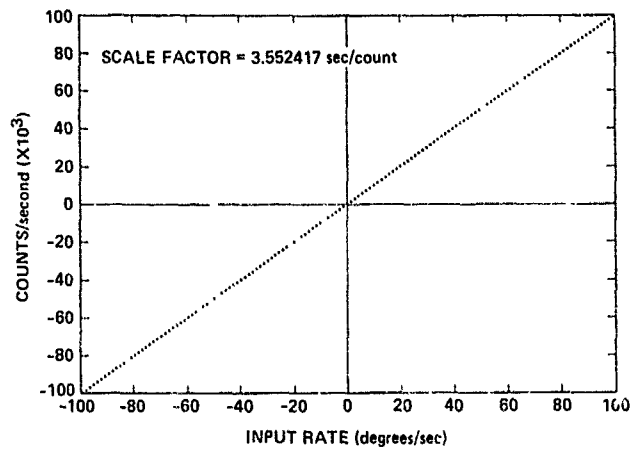


Figure 11 - Dynamic Test Data of Fiber Gyro (Uncompensated).

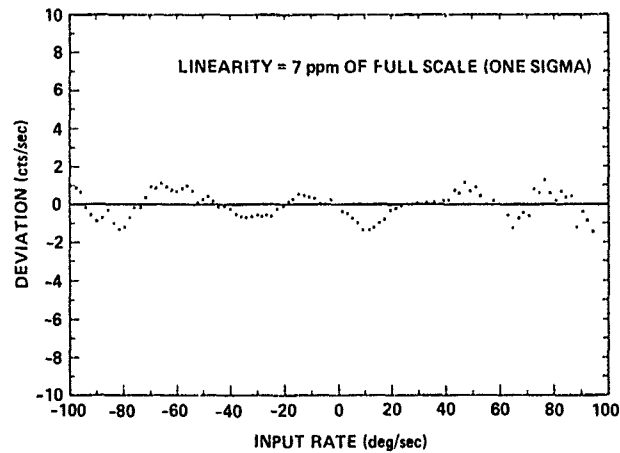


Figure 12 - Deviation of Dynamic Data from Ideal Transfer Function.

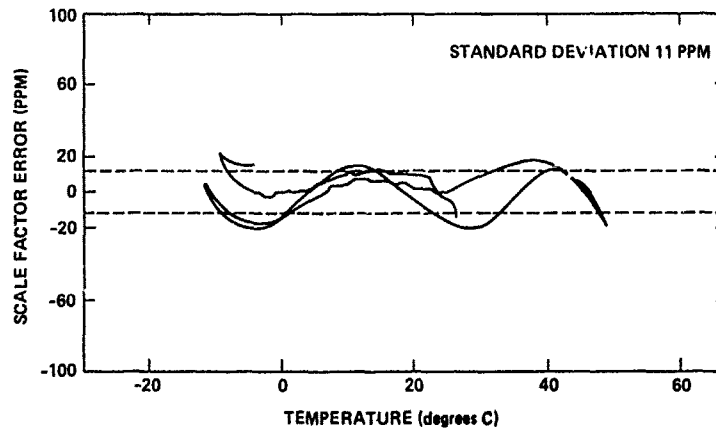


Figure 13 - Wavelength Induced Scale Factor Error.

VI. CONCLUSION

Significant advances have occurred in fiber-optic gyros over the last twelve years. Sensitivity to rotation has improved by almost 7 orders of magnitude since 1976. Sources of error have been identified and reduced. Current development activities are directed towards improving scale factor stability and linearity for strapdown applications, improving environmental ruggedness including radiation hardness, reliability, manufacturability, and design to cost.

Meeting the performance requirements of modern inertial navigation requires that each of the many sources of error be properly understood and controlled. Careful design is essential to achieving this. However, development gyros meeting many of the requirements for inertial navigation have been built. In their present forms they hold the promise of a real solution to the ever present problem of getting higher performance at lower cost.

REFERENCES

- ¹Sagnac, G., "L'ether lumineux demontre par l'effet du vent relatif d'ether dans un interferometre en rotation uniforme," C.R. Acad. Sci. (Paris) 157 (1913), p. 705.
- ²Michelson, A. A., and Gale, H. G., Nature, 115 (1925), p. 566.
- ³Vali, V., and Shorthill, R. W., "Fiber Ring Interferometer," Appl. Opt. 15 (1976), p. 1099.
- ⁴Arditty, H., Shaw, H. J., Chodorow, M., and Kompfner, R., "Re-entrant Fiberoptic Approach to Rotation Sensing," Proc. Soc. Photo-Opt. Instrum. Eng., 157 (1978), p. 138.
- ⁵Davis, J. L., and Ezekiel, S., "Techniques for Shot-Noise-Limited Inertial Rotation Measurement Using a Multiturn Fiber Sagnac Interferometer," Proc. Soc. Photo-Opt. Instrum. Eng., 157 (1978), p. 131.
- ⁶Cahill, R. F., and Udd, E., Opt. Lett., 4 (1979), p. 93.
- ⁷Ulrich, R., "Fiber-Optic Rotation Sensing With Low Drift," Opt. Lett., 5 (1980), p. 173.
- ⁸Cutler, C. C., Newton, S. A., and Shaw, H. J., "Limitations of Rotation Sensing by Scattering," Opt. Lett., 5 (1980), p. 488.
- ⁹Bergh, R. A., Lefevre, H. C., and Shaw, H. J., "All-Single-Mode Fiber-Optic Gyroscope," Opt. Lett., 6 (1981), p. 198.
- ¹⁰Lefevre, H. C., Bergh, R. A., and Shaw, H. J., "All-Single-Mode Fiber-Optic Gyroscope With Long Term Stability," Opt. Lett., 6 (1982), p. 502.
- ¹¹Davis, J. L., and Ezekiel, S., "Closed-Loop, Low-Noise Fiber Optic Rotation Sensor," Opt. Lett., 6 (1982), p. 505.
- ¹²Ezekiel, S., Davis, J. L., and Hellwarth, R. W., "Intensity Dependent Non-Reciprocal Phase Shift in a Fiberoptic Gyroscope," presented at the First International Conference on Fiber Optic Rotation Sensors and Related Technologies, MIT, Cambridge, Mass., Nov. 1981, published in Fiber Optic Rotation Sensors, Springer Verlag (1982), p. 332.
- ¹³Bergh, R. A., Culshaw, B., Cutler, C. C., Lefevre, H. C., and Shaw, H. J., "Source Statistics and the Kerr Effect in Fiber-Optic Gyroscopes," Opt. Lett., 7 (1982), p. 563.
- ¹⁴Lefevre, H. C., Bergh, R. A., and Shaw, H. J., "All-Fiber Gyroscope with Inertial-Navigation Short-Term Sensitivity," Opt. Lett., 7 (1982), p. 454.
- ¹⁵Burns, W. K., Moeller, R. P., Villarruel, C. A., and Abebe, M., "Fiber Optic Gyroscope with Polarization Holding Fiber," presented at Topical Meeting on Optical Fiber Communication, Post Deadline Paper PD#2, New Orleans, February (1983).
- ¹⁶Arditty, H., Lefevre, H., and Graindorge, "A Fiber gyro prototype capable of meeting the full tactical grade requirements," Proc. SPIE, Vol. 468: Fiber Optics '84, 1984.
- ¹⁷Arditty, H., "Modulation and Signal Processing for the fiber optic gyro," Proc. SPIE, Vol. 719: Fiber Optic Gyros: 10th Anniversary Conference, 1986.
- ¹⁸Lefevre, H., Vatoux, S., Papuchon, M., and Puech, C., "Integrated Optics - A practical solution for the fiber optic gyro," Proc. SPIE, Vol. 719: Fiber Optic Gyros: 10th Anniversary Conference, 1986.
- ¹⁹Post, E. T., "Sagnac Effect," Rev. Mod. Phys., 39 (1967), p. 475.
- ²⁰Arditty, H. J., and Lefevre, H. C., "Sagnac Effect in Fiber Gyroscopes," Opt. Lett., 6 (1981), p. 401.

- ²¹Arditty, H. J., and Lefevre, H. C., "Electromagnetisme des Milieux Dielectriques Lineaires en Rotation et Application a la Propagation D'ondes Guidees," Appl. Opt., 21 (1982), p. 1400.
- ²²Carroll, R., Coccoli, J., Cardarelli, D., Coate, G., and Draper, C., "Passive Fiber Optic Resonant Ring Gyro and Comparison with Interferometer Gyro Approaches," Proc. SPIE, Vol. 719: Fiber Optic Gyros: 10th Anniversary Conference, 1986.
- ²³Iwatsaki, K., Hotate, K., and Higashiguchi, U., "Effect of Rayleigh Back Scattering in an optical passive ring-resonator gyro," Applied Optics, Vol. 23, No. 21, 1 November 1984.
- ²⁴Sheem, S. K., and Giallorenzi, T. G., "Single-mode Fiber-optical Power Divider: Encapsulated Etching Technique," Opt. Lett., 4 (1979), p. 29.
- ²⁵Bergh, R. A., Kotler, G., and Shaw, H. J., "Single-Mode Fiber-Optic Directional Coupler," Electron. Lett., 4 (1979), p. 29.
- ²⁶Burns, W. K., Moeller, R. P., Villarruel, C. A., and Abebe, M., "Fiber-Optic Gyroscope with Polarization-Holding Fiber," Opt. Lett., 8 (1983), p. 540.
- ²⁷Bohm, K., Marten, P., Petermann, K., and Weidel, E., "Low-Drift Fiber Gyro Using Super Luminescent Diode," Electron. Lett., 17 (1981), p. 352.
- ²⁸Ulrich, R., and Johnsin, M., "Fiber-Ring Interferometer: Polarization Analysis," Opt. Lett., 4 (1979), p. 152.
- ²⁹Kintner, E. C., "Polarization Control in Optical-Fiber Gyroscopes," Opt. Lett., 6 (1981), p. 154.
- ³⁰Schupe, D. M., "Thermally Induced Non-Reciprocity in the Fiber-Optic Interferometer," Appl. Opt., 19 (1980), p. 654.
- ³¹Frigo, N. T., "Compensation of Linear Sources of Nonreciprocity in Sagnac Interferometers," Proc. Soc. Photo-Opt. Instrum. Eng., 412 (1983), p. 268.
- ³²Bohm, K., Petermann, K., and Weidel, E., "Sensitivity of a Fiber-Optic Gyroscope to Environmental Magnetic Fields," Opt. Lett., 7 (1982), p. 180.
- ³³Hotate, K., and Tabe, K., "Drift of an optical fiber gyro caused by the Faraday effect: Influence of the Earth's magnetic field," Appl. Optics, Vol. 25, No. 7, 1 Apr 1986.
- ³⁴Ezekiel, S., Davis, J. L., and Hellwarth, R. W., "Observation of Intensity-Induced Nonreciprocity in a Fiber-Optic Gyroscope," Opt. Lett., 7 (1982), p. 457.
- ³⁵Petermann, K., "Intensity-Dependent Nonreciprocal Phase Shift in Fiber-Optic Gyroscopes for Light Sources with Low Coherence," Opt. Lett., 7 (1982), p. 623.
- ³⁶Frigo, N. J., Taylor, H. F., Goldberg, L., Weller, J. F., and Rasleigh, S. C., "Optical Kerr Effect in Fiber Gyroscopes: Effects of Non-Monochromatic Sources," Opt. Lett., 8 (1983), p. 119.
- ³⁷Bergh, R. A., Lefevre, H. C., and Shaw, H. J., "Compensation of the Optical Kerr Effect in Fiber-Optic Gyroscopes," Opt. Lett., 7 (1982), p. 282.
- ³⁸Udd, E., and Cahill, R. F., "Compact Fiber-Optic Gyroscope," presented at the First International Conference on Fiber-Optic Rotation Sensors and Related Technologies, MIT, Cambridge, Mass., Nov. 1981, published in Fiber-Optic Rotation Sensors, Springer Verlag (1982), p. 302.
- ³⁹Bailly, A. J., Dye, M. S., and Traynen, B. T., "Dispersion-Induced Nonreciprocal Effects in Phase Nulling Fibre Gyroscopes," presented at the First International Conference on Optical Fibre Sensors, London, April 1983, p. 136.
- ⁴⁰Kim, B. Y., Lefevre, H. C., Bergh, R. A., and Shaw, H. J., "Harmonic Feed-Back Approach to Fiber Gyro Scale Factor Stabilization," presented at the First International Conference on Optical Fibre Sensors, London, April 1983, p. 136.
- ⁴¹Kim, B. Y., and Shaw, H. J., "Gated Phase Modulation Feedback Approach to Fiber-Optic Gyroscopes," to be published.
- ⁴²Davis, W. C., Pondrom, W. L., Thompson, D. F., "Fiberoptic Gyro Using Magneto-Optic Phase Nulling Feedback," presented at the First International Conference on Fiber-Optic Rotation Sensors and Related Technologies, MIT, Cambridge, Mass., Nov. 1981, published in Fiber-Optic Rotation Sensors, Springer Verlag (1982), p. 308.

USE OF A THREE-AXIS MONOLITHIC RING LASER GYRO AND DIGITAL SIGNAL PROCESSOR IN AN INERTIAL SENSOR ELEMENT

by

Donald J. Weber
The Singer Company
Kearfott Guidance and Navigation Division
Little Falls, New Jersey 07424
United States

SUMMARY

Over the last decade the Ring Laser Gyro (RLG) has been introduced into various navigation applications. This technology continues to show performance growth relative to traditional "iron" gyros, while demonstrating greater reliability and lower power consumption. In order to realize the advantages of this technology, design considerations have included additional computational requirements, increased size/weight, and incorporation of mechanical gyro dither (to overcome the RLG lock-in phenomena).

Coincident with advances in RLG technology to perform at levels consistent with those required for navigation systems has been the tremendous increase in computational capability of available processors. Initial configurations utilizing RLGs required special-purpose/dedicated processors to perform the "strapdown algorithm." This differed from traditional iron gyro inertial platform configurations, where electro-mechanical gimbals maintain the physical attitude of the inertial sensors. The advances in digital signal-processor technology now permit an architecture that eliminates special-purpose processors and reduces the computational load on the general-purpose processor in the Inertial Navigation System (INS).

The other issue of size and weight for a specified performance range is now yielding before the next generation of RLGs. In this vein, Kearfott has developed a TRIPLE Laser Gyro (TRILAG™). The TRILAG is a three-orthogonal-axis, monolithic RLG based on a design of nested optical cavities in which the mirrors are shared between the optical cavities. The six mirrors (visualized as the faces of a cube) form three square orthogonal gyros with each mirror shared by two gyros. This integrated approach provides the largest possible RLG path length in a given volume, provides increased alignment stability, and provides for a reduction in parts count, which lowers manufacturing costs while increasing reliability. The current outlook indicates that inertial systems configured around this next-generation sensor will find application in terrestrial, aircraft, and missile configurations. This forecast is now supported by laboratory and flight testing at both gyro and system level.

This paper presents a configuration based upon a 24-cm path length TRILAG gyro, three miniature single-axis force rebalance accelerometers, and a digital signal processor. The resulting inertial sensor element provides navigation performance equal to a conventional medium-accuracy navigation system (0.4 to 1 nmi/h) with half the weight and volume of a traditional RLG configuration. Its application as a stand-alone inertial subsystem or as an integral part of a full inertial navigation system will be discussed and performance data presented. Companion development utilizing down-sized TRILAG gyros for tactical missile or sensor stabilization applications is also reviewed.

1.0 INERTIAL SENSOR ELEMENT EVOLUTION

Over the past three decades, the INS has evolved as sensors and support electronics have increased in capability and decreased in size. The medium-accuracy (1 nmi/h) aircraft INS, which is aligned in about 10 minutes (versus a 1 hr or 30 s alignment system) and operates for 1 to 2 hr, is a major application. The rotating-wheel iron gyros developed in the early 1950s initially provided this capability. At that time, Inertial Sensor Elements, ISEs (including gyros, accelerometers, a gimbal set, and dedicated support electronics), weighed 20 lb and occupied 300 in³. These were combined with first-generation airborne digital computers that required (nominally) five cards of electronics (5 lb) and a core memory that contained 8 K (8192) or 16 K words of memory (another 5 lb). Throughput of 10⁵ fixed-point operations per second was typical. With power supply, input/output electronics, and chassis, the INS weighed 75 lb and was a full ATR size.

Evolution into the 1970s reduced the iron gyro ISE to 10 lb and 150 in³. The associated airborne computer evolved to one card of electronics and one card of solid-state memory (2 lb total). The memory contained 64 K words, and throughput of 5 x 10⁵ operations per second was available. Arithmetic capability in floating point had been introduced.

Development in the last decade has been directed at performance and reliability enhancements. Through evolutionary (or revolutionary) change in design, the iron gyro ISE can now be manufactured with precision accuracy (0.1 to 0.2 mi/h) with extended alignment times (30 minutes) in the same form factor. The associated digital computer memory now contains 256 K words, and throughput is measured in MFLOPS (millions of

floating point operations per second). The INS has been reduced to 30 lb and a 3/4 ATR short form factor (U.S. Air Force Standard Inertial). The Kearfott-produced U.S. Air Force Precision Standard INS ISE is shown in Figure 1. In the 1970s, RLG technology emerged. Over the last decade it has taken its place alongside the iron gyro in the medium-accuracy aircraft marketplace. Both the Department of Defense and the commercial market are presently procuring both technologies. The first-generation RLG ISE is in the 20 lb and 300 in³ range, which results in an INS weight of 50 lb. The form factor has not changed as it is now application rather than technology driven. The Kearfott-produced U.S. Navy CAINS II RLG ISE is also shown in Figure 1.

Activity is now well advanced toward developing the second-generation RLG ISE, shown in the center of Figure 1, which reduces the weight and volume to that of the iron gyro. Development of the gyro for this ISE, TRILAG, had been initiated in 1981.

2.0 TRILAG DEVELOPMENT

The Kearfott TRILAG gyro is a monolithic three-axis RLG that uses a single block and six shared mirrors. Although it is novel in geometry, it is based on the same technology that has been successfully applied to the single-axis RLG. The TRILAG gyro is illustrated in Figure 2 and shown schematically in Figure 3. The geometry can be visualized as having six mirrors configured on the faces of a cube to form three orthogonal, conventional square RLGs. Its great structural and interaxis stability is derived from its compactness and single-block design. The gyro's reliability is enhanced, over other gyro triad designs, because the critical parts count is the smallest possible (a single block, six mirrors, one cathode, one dither mechanism, and one High-Voltage Power Supply (HVPS)).

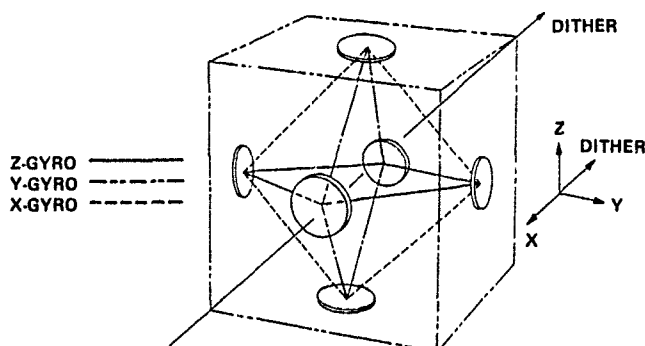


FIGURE 3. TRILAG GYRO SCHEMATIC

Concern over the beam interaction in the TRILAG gyro arises because the major RLG error source (back scattered light-induced mode locking) may serve to produce interaxis mode locking. This concern is unwarranted both theoretically and in the laboratory. Lock-in in the single-axis gyro occurs because the oppositely directed traveling waves are matched in absolute frequency to a very high degree, and thus, phase coupling is a long-term coherent process. A carefully controlled experiment, using a TRILAG gyro block, showed no pulling for even a very stringently controlled matching of absolute frequencies of two orthogonal beams.

The simultaneous alignment of the three square cavities sharing mirrors is easily accomplished. The optimal positions of the spherical mirrors are found by adjusting their position to achieve minimum loss on all three axes. Total alignment time for a three-axis TRILAG gyro is comparable to that for two single-axis RLGs. The discharge path involves six anodes and a single internal cathode designed to be symmetric so as to prevent Langmuir flow bias effects. All seals are made either by optical contact (mirrors) or by the highly reliable indium step-seal method pioneered at Kearfott (electrodes, getter, and fill tube). These design features contribute to long life and high reliability.

3.0 TRILAG ISE/INS DEVELOPMENT

Following successful testing of the TRILAG concept, activity was initiated to incorporate this new sensor into an ISE and then into an INS for test, evaluation, and demonstration. The ISE utilizing the TRILAG to provide three axes of attitude data includes three Kearfott single-axis MOD VII force rebalance accelerometers. The RLG lock-in phenomenon is suppressed in the TRILAG ISE by mechanically dithering the block along a single axis. This provides a single frequency dither for all three axes. The dither mechanism in the TRILAG gyro consists of two dither hinges with spokes driven by means of Piezoelectric Transducers (PZTs). The two hinges deliver the same torque about the dither axis and, being equally stiff, provide an inherently symmetrical dither action. This minimizes potential coning/rocking problems.

A development model INS utilizing three single-axis RLGs configured with a single dither was modified for test of the prototype TRILAG. As discussed above, other than mechanical modifications to mount the new gyro and accelerometer triad, only the path-length control electronics required minor adaptation from that used with conventional single-axis RLGs.

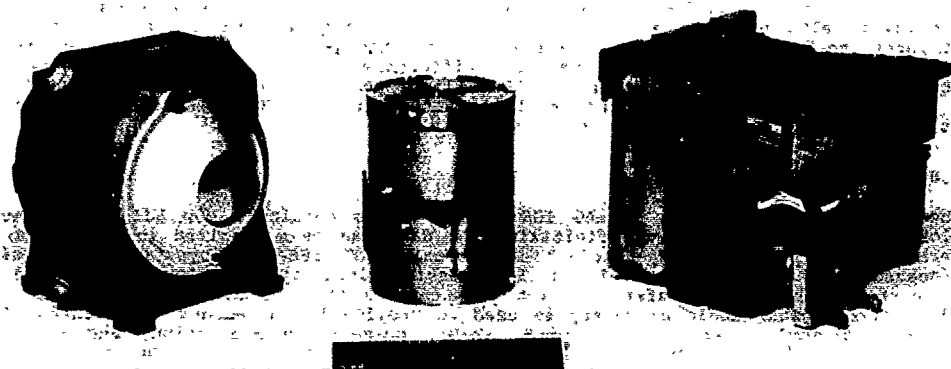


FIGURE 1. MEDIUM-ACCURACY ISE

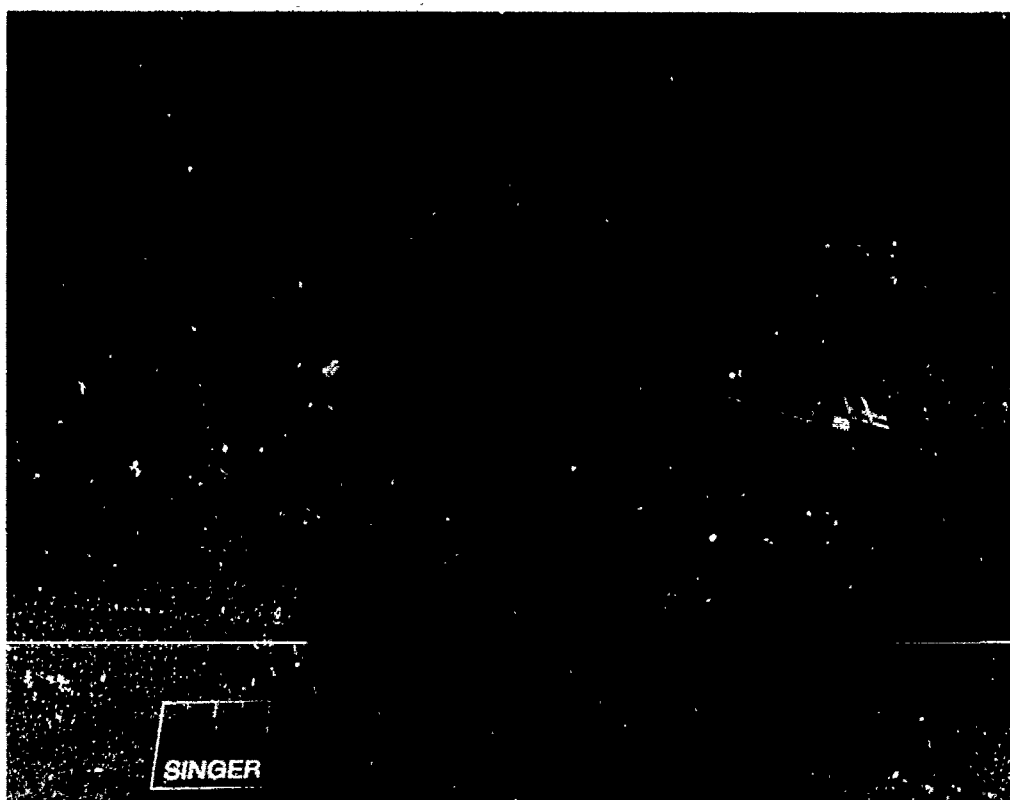


FIGURE 2. TRILAG GYRO ON FILL STAND

4.0 TRILAG FLIGHT TEST RESULTS

Early Kearfott flight testing was concluded in 1985. Since then, flight test evaluations have been conducted by Boeing Aerospace Company (BAC) and McDonnell Douglas Astronautics Company (MDAC). The unit was tested as an INS by MDAC on a series of five 90-min flights out of St. Louis, MO. The unit's performance of 1.4 nmi/h (Figure 4) during the flight test was consistent with expected results based upon preflight laboratory evaluation.

With specifically modified software to simulate an Inertial Measurement Unit (IMU) providing delta velocity and delta attitude, the unit was interfaced with a central computer that performed in-air alignment and navigation, simulating a short-range missile scenario. Position and velocity differences (with respect to the master INS) are shown in Table 1 for both BAC and MDAC flight tests. Projected position errors from a computer simulation of the MDAC test trajectory and the TRILAG error characteristics at the 10-min point were 310 ft (east and north).

5.0 DIGITAL SIGNAL PROCESSOR

The second sensor in the TRILAG ISE is the force rebalance MOD VII accelerometer. Closing the capture loops and digitizing the data can be mechanized in many ways. The mechanization employed in Kearfott single-axis RLG ISEs employs a digital signal processor. Programmable digital signal-processing circuitry is employed within the loop to perform the loop transfer function computations for all three accelerometer axes. Additional digital logic circuitry is used to provide the necessary loop control functions that include Analog to Digital (A/D), pulse-width modulation, and self-test. Figure 5 is a block diagram of the accelerometer electronics. The amplitude-modulated pickoff error signal of each accelerometer is demodulated, filtered, and then fed to a tracking A/D converter. The output of the A/D converter is sampled at a fixed rate and fed to the digital signal-processing circuitry.

A second-generation Digital Signal Processor (DSP) was incorporated into the ISE electronics. The approach is to incorporate accelerometer capture-loop processing along with gyro control functions and high-speed accelerometer/gyro processing. Additionally, the DSP performs orthogonalization, thermal modeling, coning, and sculling compensation for the accelerometers and gyros. The data is separated into two flows. One performs high-speed filtering to remove aliasing effects for flight control applications. The other flow performs the coning and sculling compensations necessary for inertial navigation. The objective is to develop, in the IMU package, compensated body referenced velocity and rotational information. Provisions are made to allow high-speed filtering of the data for use in autopilot applications.

The DSP chosen was Texas Instrument's TMS320C25. This processor is a logical extension of the electronics based on the first-generation TMS320C10. The TMS320C25 is a CMOS device that has a throughput capacity of up to 10 million instructions per second (MIPS). Provisions in the hardware architecture of the processor allow it to be connected in a multiprocessor configuration. This provides for a system configuration illustrated in Figure 6. The ISE and electronics can be configured as a stand-alone IMU, combined with a general-purpose processor and input/output electronics in a conventional Line Replaceable Unit (LRU) INS, or organized into a Line Replaceable Module (LRM) consistent with the U.S. Air Force Pave Pillar configuration.

6.0 ALTERNATIVE TRILAG CONFIGURATIONS

The core electronics and TRILAG discussed above provide the basic architecture for a family of ISEs that promise to eliminate the weight penalty now associated with an RLG. A valuable variable that can be exercised at this point is the scaling capability of the RLG. For a given quality of components, mirrors in particular, as the path length (physical size) increases, the performance of the instrument improves. This permits a size-versus-performance trade-off to adapt the TRILAG to various classes of applications.

The tactical missile class performance is being addressed by a 16 cm path length TRILAG presently under development. The U.S. Navy at the Naval Weapons Center (NWC) and the U.S. Air Force at Eglin AFB are active in this application. The family is presently being expanded, with initial designs of a high-accuracy (60 cm) and a very low-performance (10 cm) gyro being undertaken.

7.0 ACKNOWLEDGMENTS

TRILAG design and technical materials were developed by the Kearfott Research Department, J. Stiles, Director; J. Simpson, Manager; B. Ljung; and J. Koper.

MDAC test data was provided by H. Routburg, A. Esker, and G. Waters.

BAC test data was provided by the Boeing (Seattle) flight test department, R. Bryant, L. Hanvey, K. Riggs, and R. Suchland.

T-24 TRILAG FLIGHT TEST ACCURACY RESULTS

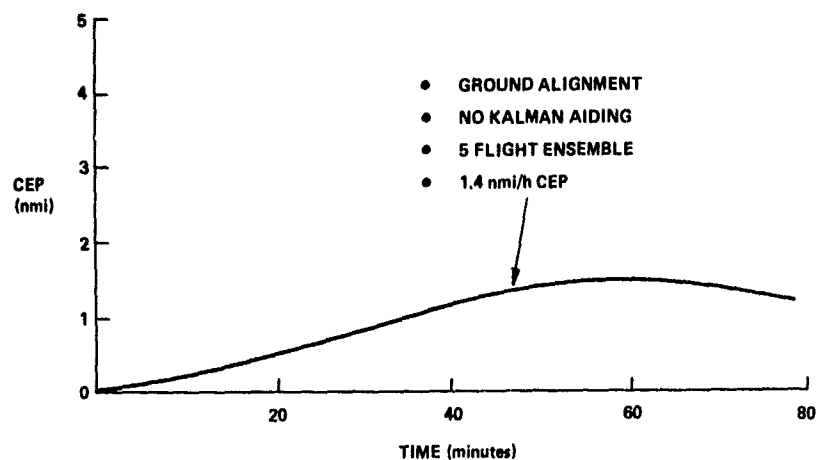


FIGURE 4. T-24 TRILAG GYRO FLIGHT TEST NAVIGATION RESULTS

TABLE 1. TRILAG GYRO FLIGHT TEST AIR ALIGNMENT RESULTS

<ul style="list-style-type: none"> • FREE INERTIAL ERRORS FOLLOWING IN AIR ALIGNMENT <ul style="list-style-type: none"> • REFERENCED TO PURE INERTIAL MASTER INS • MDAC AND BOEING TEST AIRCRAFT • 1 SIGMA OF ALL RUNS 				
	POSITION ERROR (ft)		VELOCITY ERROR (ft/s)	
	EAST	NORTH	EAST	NORTH
5 MINUTES				
MDAC	68	51	0.5	0.4
BOEING	132	72	1.1	1.2
10 MINUTES				
	311	250	1.2	1.1
	505	450	2.8	3.3

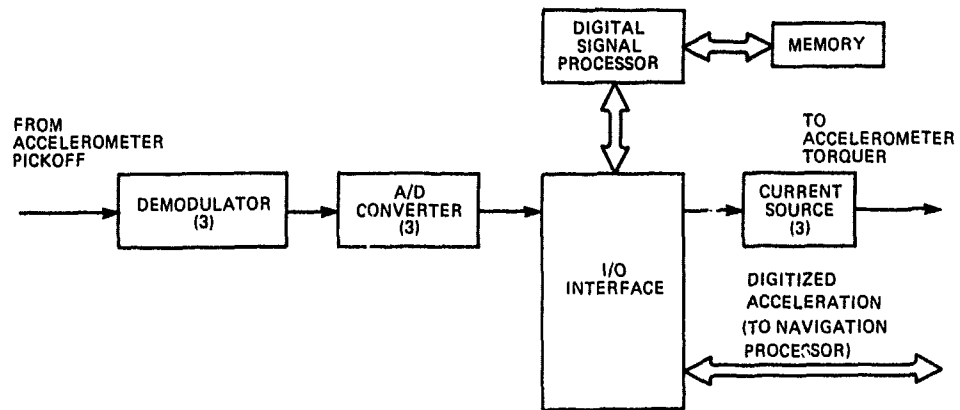


FIGURE 5. PROCESSOR-CONTROLLED ACCELEROMETER LOOPS

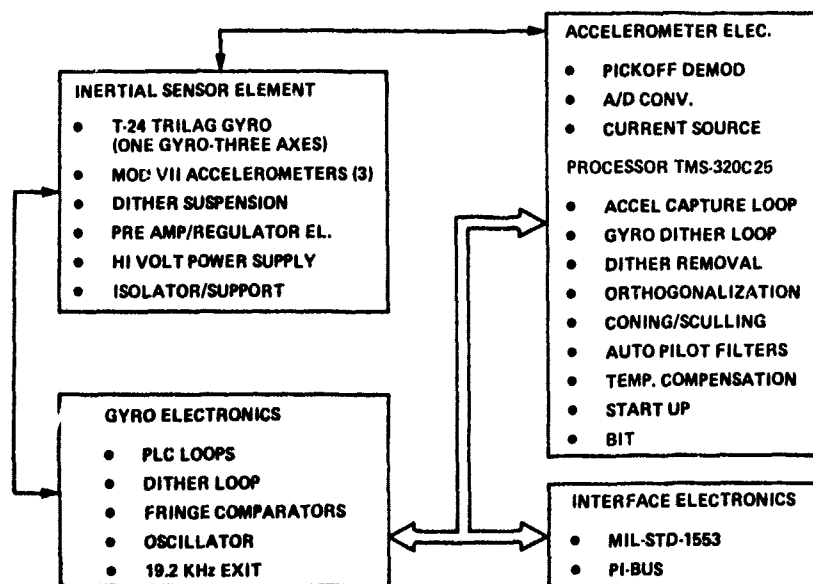


FIGURE 6. TRILAG GYRO IMU FUNCTIONAL BLOCK DIAGRAM

RING LASER GYRO MARINE INERTIAL NAVIGATION SYSTEMS

by

Clinton C. Remuzzi
Manager Marine Systems Engineering
Litton Systems Inc.
Guidance and Control Systems Division
5500 Canoga Avenue
Woodland Hills, California 91367-6698
United States

SUMMARY

The NATO navies currently use gimballed spinning wheel gyros to support their inertial navigation system requirements. Ring Laser Gyros (RLGs) are now being considered for the next generation of marine inertial navigators. RLGs are expected to provide better performance (longer extrapolation intervals between resets), higher reliability, and lower cost-of-ownership.

The performance improvement expected from RLGs had already been demonstrated at sea during prototype system tests conducted by the U.S. Navy. The final production design of an RLG marine inertial navigation system (INS) required that tradeoffs be made among the sometimes conflicting factors of performance, reliability and environmental isolation.

This paper describes some of the considerations and tradeoffs involved in the design of the Litton marine RLG INS. It is hoped that this information will prove useful for those interested in the field both as designers and as users.

INTRODUCTION

The use of inertial navigators aboard surface ships has expanded during the 1980's, at least in part, as the result of the introduction by Litton in 1978 of the WSN-5 system. This system, designed to meet Navy specifications, is an accurate, compact, reliable equipment which is cost effective enough to allow wide deployment in the surface fleet. Litton also developed a gyrocompass during this period, the WSN-2, which is also the U.S. Navy standard. An LSN navigator, which is being sold to other navies, is the final member of this family of marine inertial products.

In order to respond to a perceived need for improved accuracy, higher reliability, and lower cost-of-ownership, Litton began a company-funded IRAD program in 1984 to develop an RLG version of the WSN-5 IMU. A very important goal of this program was to retain as much of the proven WSN-5 hardware and software as possible. Since Litton had already supplied over 1000 Navy standard systems, the new RLG design had to be completely compatible with the mechanical and electrical characteristics of the standard system. Furthermore, it was considered mandatory that the unit be retrofittable into existing shipboard installations without disturbing the enclosure mounting.

Various ways of mechanizing a marine RLG system were carefully examined to assure that the one selected was optimum for marine use. Design tradeoffs were made and technical issues addressed to achieve the necessary commonality, performance environmental isolation, and reliability/maintainability. These tradeoffs are described in some detail in this paper together with discussions of performance testing, design for the qual environment, and reliability/maintainability considerations.

COMMONALITY

The three basic systems which embody the Litton commonality concept are shown in Figure 1. These are: the WSN-2 - the standard U.S. Navy stabilized gyrocompass also supplied to other navies, LSN - Litton Ship's Navigator primarily designed for export and the WSN-5 - the standard U.S. Navy surface ship inertial navigator. The WSN-5 is further described in reference 1 and a detailed discussion of the common elements among the three systems is contained in reference 2. The fact that more than 1000 of these systems have been sold constrained the type of changes considered in going from the currently used gimballed IMU to an RLG IMU. Litton marine inertial systems had already been integrated with other larger shipboard systems. Litton perceived that users would like to enhance performance and reliability using an RLG IMU without having to install a totally new system, or even removing existing enclosures from the ships. For these reasons it was considered mandatory that the new RLG IMU fit inside the existing enclosure and also be field retrofittable without removing the enclosure from the ship.

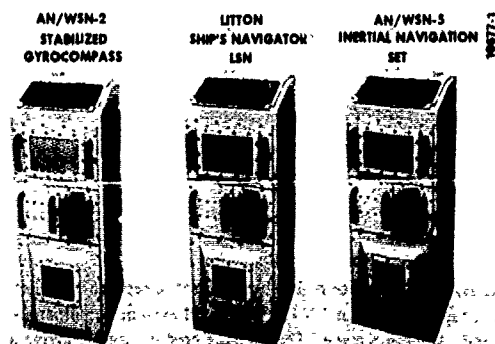


Figure 1. Litton's Three Basic Marine Systems

The major assemblies of the RLG WSN-5 are shown in Figure 2. The significant changes to the present configuration are as follows:

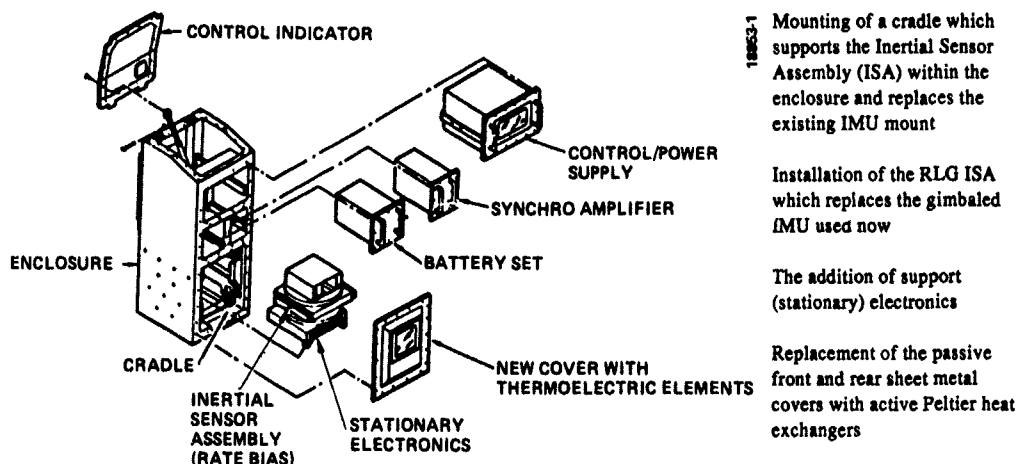


Figure 2. RLG WSN-5 Major Assemblies

Exchange of two processor boards within the Control/Power Supply (CPS) to implement the Synchronous Data Link Control (SDLC) bus

Reprogramming (replacement) of the PROM card in the CPS drawer with the new software required for the RLG.

Steps actually required to accomplish a field modification of a WSN-5 system are listed below to demonstrate the ease with which the change can be accomplished.

1. Remove the lower enclosure covers (front and back), the gimballed IMU, and the IMU supporting structure, and the precision restoring devices (PRDs).
2. Install the circulator fan, ducting and insulating panels.
3. Drill additional holes in the side of the enclosure and mount the cradle used to support the ISA.
4. Install the harness provided with the RLG modification kit. Change the termination on ten wires and add six wires to the main harness. Add one wire to the Control/Power supply drawer. Mount the Peltier cover connectors.
5. Remove the I/O #2 and #3 cards from the CPS drawer and replace with the new cards. Install the reprogrammed PROM card.
6. Install the stationary electronics card rack and three electronics cards.
7. Align the cradle pads to the ship's attitude references using the ISA alignment fixture.
8. Install the ISA in the cradle.
9. Mount the front and rear Peltier coolers.

The system is now ready to turn on and operate as an RLG navigator!

HARDWARE/MECHANIZATION TRADEOFFS

In order to upgrade the Litton marine navigators with RLGs, an RLG sensor had to be designed to replace the existing gimballed IMU. When the company funded IRAD program began at the end of 1984, a number of system implementations were examined. Serious consideration was given to each and a selection made. The tradeoff process used to arrive at the final ISA design is discussed in the following "design decisions" section.

Design Decisions

The first major decision was to select a rate bias approach to provide gyro "unlocking" in preference to dithering. Litton had demonstrated a dithered RLG system very successfully in a U.S. Navy test aboard the Vanguard. Furthermore Litton was (and is) producing dithered gyros in substantial numbers for commercial and military aircraft navigators. Nevertheless it was decided not to use dithering in the marine system for reasons discussed in the subsequent paragraphs.

Systems using dithered gyros have three major drawbacks in the marine application. These are:

Dithered gyros generate large amounts of structure borne (and airborne) noise at the dither frequency(ies). This acoustic noise is only a problem for ships where the sound traveling through the water can be detected. Other vehicles, such as aircraft, have no problems with acoustic noise. The dither generated noise can be mitigated to some extent by counterbalancing, i.e., shaking the gyro mounting block with an equal and opposite force. This mechanization can become rather complex and costly if the gyros operate at different frequencies as they conventionally do. Mechanical isolation or filtering can also be used but must be carefully designed to avoid compromising attitude accuracy. Mechanical filters can also present problems when attempting to design for qualification shock and vibration levels.

Dithered gyros tend to have unacceptably large shifts in axis alignment when subjected to the full marine qualification shock levels. The dither mechanism by its very nature is a compromise. Ideally, it would have infinite compliance about the dither axis and infinite stiffness about the two orthogonal axes. Since this is impossible, the second most desirable characteristic would be perfect returnability, i.e., no deformation or shift following application and removal of a force. Again, this is not achievable in a real design, so even if shock mitigation is used, it is difficult if not impossible to assure stability of alignments. Stability must be maintained during and immediately after the application of a MIL-S-901C hammer blow. If stability is not maintained, vehicle motions will produce the equivalent of gyro drifts. Calibration following such alignment shifts is not a practical strategy because the navy specifications generally require full performance within a short period following the shock application.

A given gyro "physics package" ceramic frame, mirrors, etc., will have 2 to 4 times more angle random walk when dithering than it will operating in the rate bias mode. Rate bias allows the gyro to operate close to the "quantum limit." Techniques for improving the performance of dithered gyros by reducing the noise generated as the gyro passes through lock-in are being developed but they involve additional hardware and software. The rate bias approach still provides the lowest random noise for a given gyro package.

The rate bias mechanization requires only a single drive versus the three equivalent drives using dither. This potential for enhanced reliability is an additional advantage of rate bias but was not a prime consideration in its selection.

The next technical tradeoff addressed the selection of a reversing ("Maytag") action in preference to continuous rotation in one direction as a means of generating rate bias. This selection was made primarily to maximize reliability. Continuous rotation in one direction is most easily accomplished using slip rings. Obviously if slip rings are used wear must be considered. Assuming a 60 deg/sec rate and 6000 operating hours per year, 3.6 million rotations would accumulate on a rate bias slip ring per year. This potential wear problem, coupled with the fact that slip rings are one of the components which contribute to low reliability in gimballed IMUs, led to the selection of reversing action in preference to continuous rotation. Reversing action is readily implemented without slip rings. As is usual in the engineering process, there are a number of tradeoffs. Continuous rotation never goes through lock-in and therefore the gyro angle random walk is exactly at the quantum limit. A scale factor error will, on the other hand, cause an increasing error with time about the rate bias axis when continuous rotation is used. Reversing action avoids this increasing angular error but does pass through lock-in twice per cycle. The random walk component generated by passing through lock-in is made negligible by passing through it at high speed and by limiting the number of reversals per unit time to a small percentage of dither reversals. This relative infrequency of reversal (4 orders of magnitude lower) also accounts for the insignificant amount of the airborne and structureborne noise generated in a rate bias system.

The use of a second rotation axis orthogonal to the first was also considered. While the potential for improved accuracy exists with a second axis, this advantage must be traded off against size, reliability, cost and other penalties in a given application. This was done through simulation for the RLG WSN-5 navigator. The results showed that the rate bias mechanization, with a single axis of rotation, had sufficient performance margin to comfortably meet the Navy's present and anticipated requirements. For those special applications requiring very high performance a second rotation axis could be added with the consequent increase in complexity.

Having selected a single axis, rate biased, reversing action implementation, the only major decision remaining was to select the gyro size (path length). When the marine development program began, Litton was already producing the 28 cm and planned to produce two additional gyro sizes; a 40 cm and a 12 cm. The 28 cm had preference because it was in production and was expected to have the lowest cost for the foreseeable future. The error budget for random walk was then

The rate bias torque motor must drive the table at constant velocity in one direction for two revolutions and then rapidly reverse. This motion would cause rapid commutator wear in a conventional torque motor. The solution to this potential problem was to use a brushless torque motor which employs electronic switches driven by Hall effect devices for commutation. This eliminates mechanical contact and consequent wear.

A transducer is necessary in a rate bias system to measure the relative angular position between the instrument cluster and the mount. Again wear and reliability considerations dictate that this be done without "touching." This was achieved in the present design using an Inductosyn which also provides advantages of size and cost over an encoder of equivalent resolution and accuracy.

The technical issues described in this section were solved in such a way as to minimize wear and maximize reliability. These characteristics are desirable in any inertial system but are particularly important in a marine system used on navy vessels with extended mission times and long supply lines. In order to verify the system concepts described here the design was completed and an engineering prototype unit built and tested. The principle mechanical features of the design are shown in Figure 4. The actual ISA hardware is shown in Figure 5.

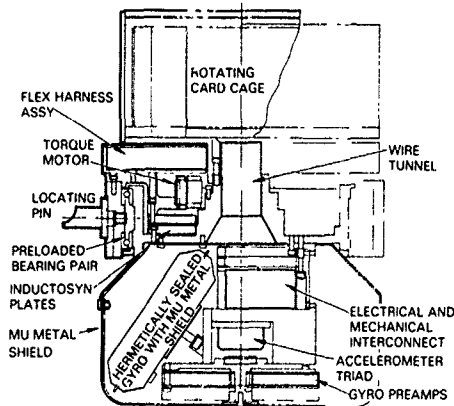


Figure 4. Marine RLG Inertial Sensor Assembly



Figure 5. Marine RLG Inertial Sensor Assembly (ISA)

PERFORMANCE TESTING

There are three levels of performance normally used to evaluate marine inertial navigators under nominal environmental conditions:

- Static
- Dynamic
- Shipboard

Static as the name implies involves running the system in a stationary lab environment and measuring the errors in attitude, velocities, and position as a function of time. Dynamic testing is also performed in the lab using two types of motion; turntable and scorsby. The turntable test requires that the system be rotated about a vertical axis usually to cardinal headings with data taken at each heading for one or more hours. This test generally has larger errors than the static test because changes in gyro drift and/or accelerometer nulls are produced as result of changing magnetic, thermal or kinematic environments with system rotation. The scorsby test involves simultaneous nonsynchronous (for marine) motion about the heading, roll, and pitch axes which simulates the motion of a vessel at sea. Scorsby tests are generally type (first article) tests and are not usually used as part of production testing.

Shipboard performance is normally worse than lab performance because other error sources which do not exist in the lab come into play. These include the effects of ocean currents, log errors, gravity anomalies and vehicle linear and rotational accelerations. The customer generally specifies performance at sea. The supplier must subtract (on an rss basis) the predicted at-sea errors to arrive at the factory sell-off specification. This usually consists of some combination of the static and dynamic lab tests already described.

The RLG WSN-5 is designed to have enough performance margin so that it "easily" passes factory sell-off with gyros and accelerometers whose performance is close to the mean for their production run.

A prototype RLG WSN-5 system was assembled by installing the various RLG components into a WSN-5 enclosure. This system was then subjected to the static and dynamic tests described in the preceding paragraphs. All runs are made following the alignment time allowed by the U.S. Navy classified specification. The results of these tests are classified, but in general, the performance goals specified by the U.S. Navy are being met with a comfortable margin. Experience has shown that any system which performs well in the lab will meet its at-sea requirement; allowing for the predicted additional errors mentioned before. The errors attributable to the rate bias mechanization while present, have negligible effect on system performance. The ISA is calibrated in the lab using a procedure similar to that described in reference 3. Since the ISA is inherently extremely stable, on the calibration terms found during the cal procedure are valid indefinitely. Other terms which, for example, vary from turn on to turn on are easily determined during the standard alignment time and can be "reset" prior to an actual performance run.

Plans are in place to evaluate the RLG WSN-5 at sea in the very near future.

ENVIRONMENTAL ISOLATION

The marine inertial system must perform within the specification before, during, and immediately after the application of a qual environmental change, i.e., the application of a MIL-S-901C hammer blow. The various steps taken to protect the RLG IMU from the harmful effects of environmental changes will be described briefly.

Thermal

The specification requires in-specification performance from 40° to 120°F. This requirement is provided for in the design using Peltier heat exchangers illustrated in Figure 6. These, acting with the other components controlling the air around the ISA, hold the air temperature to well within 1°F over the whole range of external ambients. The thermal control system is illustrated in Figure 7.

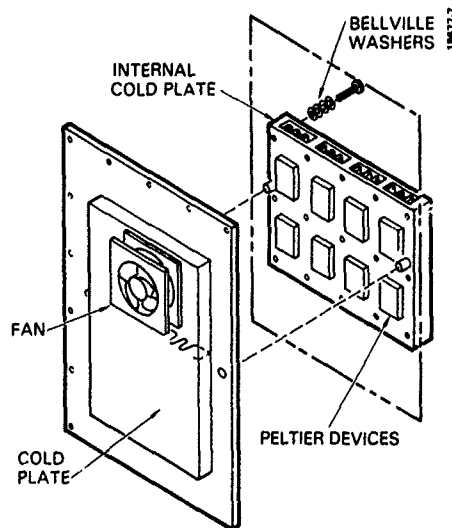


Figure 6. Thermal Control System Panel Assembly

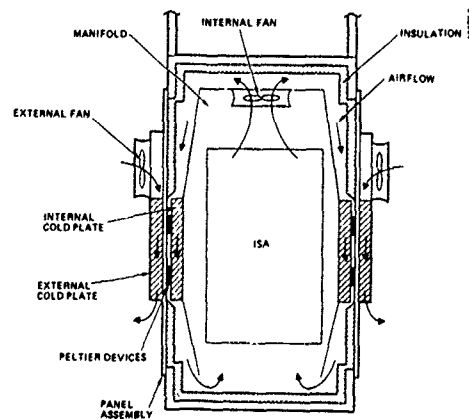


Figure 7. Thermal Control System

Magnetic

The specification for magnetic fields requires full performance up to 5 gauss. This is achieved in the ISA by double shielding. Each gyro has an individual shield and the lower ISA cover is also a magnetic shield. With these shields in place the magnetic sensitivity of the instruments is reduced to negligible values.

Shock

Since the RLG ISA uses rigidly mounted (non-dithered) gyros shock protection is easily accomplished. MIL-S-901C specifies a test procedure not a specific acceleration profile. A nominal value of 300 g - 11 millisecond is usually assumed but analysis has shown that components up to 1000 g exist at certain frequencies. The ISA mounts in a cradle which is attached to the WSN-5 enclosure using 8 vibration isolators. These isolators limit the acceleration under shock to less than 200 g. This system also provides for returnability of less than 5 arcseconds.

Other Environmental Requirements

All other environmental specs were considered in the design of this equipment. Some, like the requirement that the enclosure be drip proof, are inherent in the WSN-5 design. Others, like EMI, were taken into account in the newly designed units.

RELIABILITY/MAINTAINABILITY

The WSN-5 has a reliability, based on field data of 3008 hours MTBF versus a specification requirement of 2800 hours.

Approximately half of the failures are attributable to problems in the IMU. If a significant improvement could be made in the IMU MTBF then the overall system reliability would improve. The RLG IMU will provide the necessary improvement. A formal RLG IMU reliability analysis has not been completed yet but initial estimates are very encouraging. The estimate for the overall RLG IMU, including the ISA and the stationary electronics, is 10,638 hours MTBF. This would improve the RLG WSN-5 MTBF to 4,241 hours.

The design of the RLG ISA is such that the calibration of the instrument package is totally independent of the electronic cards. This means that the instrument package, which consists of the three gyros, the accelerometer triad, the torque motor, Inductosyn, flex harness and bearings can be considered a replaceable subassembly. The MTBF for this subassembly is estimated at 17,857 hours. Assuming the system is operated 6000 hours per year (per navy specification) this subassembly would be replaced every three years. The failed package would be returned to the repair facility for replacement of the defective component, recalibrated, and returned to the supply system.

If the gyros in the ISA are assumed to have an MTBF of 100,000 hours and the system is used 6,000 hours per year, a gyro would fail every 5 1/2 years in a given system. If an additional rotation axis were designed into the ISA then a gyro could be carried as a spare and calibrated in place without requiring replacement of the whole instrument package. Now is the components which supply the additional rotation are assumed to have a failure rate of 20 per million hours (which is a reasonable assumption) then this subsystem can be expected to fail in 8 1/3 years. This means that the user only gets the change to calibrate one gyro before the components added to provide calibration fail themselves! This analysis shows that an additional rotational or indexing axis does not effectively improve reliability/maintainability. The only justification for the additional axis is to improve performance. Since the RLG WSN-5, as previously discussed, has performance margin with a single axis, there was no rationale for including a second axis of rotation. A qualitative comparison of three classes of IMUs is given in Figure 8 which shows why the single axis rate bias mechanization offers advantages from a reliability point of view. The figure shows that the rate bias IMU has the fewest number of mechanical and/or electromechanical devices. Since these components tend to have the highest failure rates their use should be minimized to maximize reliability.

COMPONENT	RLG RATE BIAS ISA	RLG DITHERED - 2-AXIS INDEXING	GIMBALLED PLATFORM (2 D.O.F. GYROS)
TORQUE MOTOR	1	2	3
GYRO DRIVES	0	3 (DITHER MOTORS)	2 (SPIN MOTORS)
ANGLE TRANSDUCERS	1	2	3
SLIP RINGS	0	2 (MAXIMUM)	1 (MINIMUM)
BEARING PAIRS	1	3 (MINIMUM)	5 (MINIMUM)
TOTAL NUMBER OF COMPONENTS	3	10 TO 13	14 TO 17

Figure 8. Number of Mechanical/Electronic Components in Different Classes of IMUs

The RLG IMU follows the same maintain ability philosophy built into the WSN-5. Extensive hardware and software BITE is included in the newly designed ISA and stationary electronics which allows the operator to identify a malfunction to a replaceable subassembly. Replacement can be accomplished with simple hand tools without soldering. The malfunctioning subassembly can be identified even if hardware shuts the system down because the fault indicators are latching devices which continue indicate even after power has been removed. The RLG IMU will have indicators and a fault matrix similar to that shown in Figure 9 which shows the matrix for the presently used gimbaled platform.

IMU FAULT INDICATOR MATRIX					
MAINTENANCE ITEM	REFERENCE DESIGNATOR	FAULT INDICATOR			
		DS1	DS2	DS3	DS4
QUANTIZER	1A6A2A2	●	○	○	○
UPPER IMU SUBASSY	1A6A1	○	●	○	○
X-ACCELEROMETER AND ARA SUBASSY	1A6A2A1A7A15	●	●	○	○
GYRO SPIN SUBASSY	1A6A2A5	○	○	●	○
Y-ACCELEROMETER AND ARA SUBASSY	1A6A2A1A7A17	●	○	●	○
SERVO AMPLIFIER SUBASSY	1A6A2A3	○	●	●	○
UPPER GYRO SUBASSY	1A6A2A1A7MP1	●	●	●	○
LOWER IMU SUBASSY	1A6A2	●	○	●	●
LOWER GYRO SUBASSY	1A6A2A1A7MP2	○	●	●	●
Z-ACCELEROMETER AND ARA	1A6A2A1A7A19	●	●	●	●
LEGEND: ○ UNLATCHED ● LATCHED					

18677-10

Figure 9. IMU BITE Indicator Status

CONCLUSION

This paper describes the design of an RLG upgrade to the existing U.S. Navy standard surface ship navigator; the WSN-5. The change is evolutionary; it allows all the specified characteristics and proven features of the WSN-5 to be retained with enhanced performance and reliability provided by the substitution of the RLG IMU for the gimbaled platform. The RLG WSN-5 has all the necessary capabilities to become the standard surface ship navigator for the 1990's and beyond.

Acknowledgment

The design and fabrication of RLG WSN-5 was the result of the efforts of many people at Litton. Particular thanks are merited by Tony Matthews for the systems and electronic design and to Art Storjohann for the mechanical design.

REFERENCES

1. Remuzzi, C.C. & Criste, F.X. "The AN/WSN-5 - A Marine Inertial Navigator For the 1980's Surface Fleet" 15th J.S.D.E. For Inertial Systems, Philadelphia, PA. 17-19 November 1981.
2. Anthold, Robert C. "Marine Gyrocompass-To-Navigator Commonality Is The Key To An Effective Design" I.O.N. National Technical Meeting San Diego, CA. 17-19 January 1984
3. Mark, John, Tazartes, Daniel, Hilby, Timothy "Fast Orthogonal Calibration Of A Ring Laser Strapdown System" Symposium On Gyro Technology Stuttgart, Germany. 17-19 September 1986

PART IV

Integrated Communication and Navigation Systems

DISTRIBUTED CONTROL ARCHITECTURE FOR CNI PREPROCESSORS

by

V.R.Subramanyan and L.R.Stine
TRW Military Electronics and Avionics Division
One Rancho Carmel, San Diego, CA 92128
United States

Abstract

Next generation avionics systems will need to incorporate extensive integration, including the Communication, Navigation, and Identification (CNI) functions. An important element in such a highly integrated CNI system is a set of programmable preprocessors, each of which can process baseband outputs from a receiver to perform real-time signal dependent processing, such as matched filtering, PN despreading, code and carrier tracking, phase rotation, correlation, convolution, pulse shape discrimination, threshold crossing, time of arrival detection, demodulation (pulse position demodulation, DPSK, etc.), message formatting, and on-line status reporting. Control of the total integrated CNI system is characterized by a distributed-control architecture, wherein the execution times range from seconds at the data processor level down to a few nanoseconds at the preprocessor level. Any candidate control architecture for the preprocessor must support reprogrammability, flexibility in event scheduling (including manipulation of event lists to add, insert, or drop basic events), and testability, while fully meeting the requirements of each CNI function. This paper describes a distributed-control architecture for a generic CNI preprocessor that meets the above requirements.

Introduction

Recent availability of VLSI devices is making integration of CNI functions very desirable. This can lead to considerable savings in size, weight, power, and maintenance costs, while increasing availability of systems. An essential part of the integrated CNI system is the highly programmable CNI preprocessor which performs real-time waveform dependent signal processing on the baseband outputs of a receiver. Generic functions such as PN despreading, phase rotation, correlation, convolution, tracking, accurate TOA detections, threshold crossing detection, filtering, demodulation (including CCKS, DPSK, and pulse position demodulation) etc., are performed in the preprocessor. These functions place demanding requirements on any candidate CNI processor control architecture. In addition, the need for simultaneous execution of multiple CNI functions within a given CNI preprocessor is not uncommon. When one looks at the total control requirements, it becomes obvious that a centralized control structure cannot meet these demands, nor can it be built and adequately tested with any degree of confidence. Thus, any chosen control architecture must meet the throughput requirements of each active CNI subfunction within the preprocessor, while providing controllability and observability to the higher level control functions in a CNI terminal. Selecting an optimal control architecture is a trade off among hardware complexity, ease of programming, efficiency of operation, and testability requirements. This paper describes the extraordinary control demands imposed by the CNI functions, and a distributed control CNI preprocessor architecture that meets these demands in an orderly fashion, allows multiple CNI function execution, and provides controllability and observability.

Typical CNI System Architecture

A typical CNI system architecture is shown in Figure 1. A system of high speed buses interconnects the CNI preprocessors, RF groups, signal processors, and data processors. The data transfer requirements of CNI systems are characterized by short data block sizes, fast access, and rapid data transfer. The buses allow localized high-rate processing and rapid data exchange, while providing fast access. Control of such a system is highly distributed, where the level of control ranges from seconds at the DP level, to a few milliseconds at the terminal control level, and finally to a few nanoseconds at the CNI preprocessor level with the RF group performing in real time.

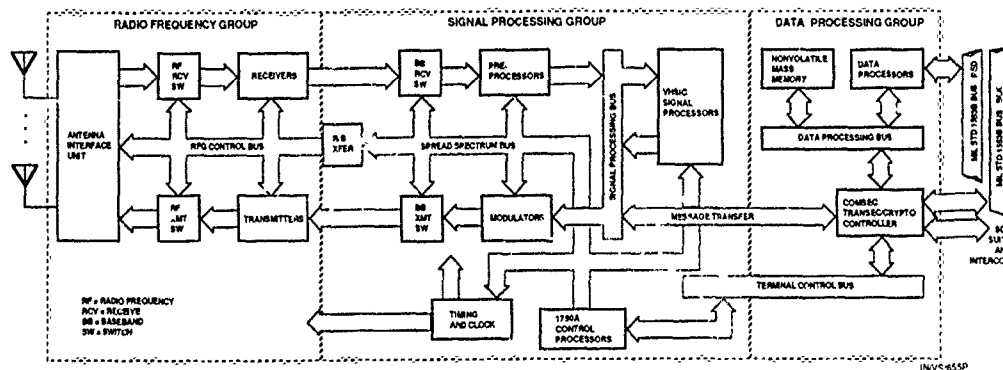


Figure 1. Typical CNI Terminal Architecture

Typical CNI Preprocessor Architecture

Figure 2 shows the autonomous processing implementation of the CNI subfunctions as elements within an integrated preprocessor. The basic architecture allows for execution of multiple CNI functions (e.g., of pairs such as JTIIDS and narrowband voice function, IFF transport and VHF Voice, and GPS and narrowband voice). The data paths through the CNI processor are reconfigurable to allow simultaneous execution of CNI functions (e.g., GPS, JTIIDS, and EJS). Note that each subfunction, like PN despreading and phase rotator, is an autonomous unit that once started, performs its function until commanded to stop. Synchronization among subfunctions to implement a CNI function is achieved by tight control of the clocking and pipelining schemes. In this paper we assume that such a timing reference, clock generation and distribution scheme are in place.

Control Requirements of CNI Functions

The control hierarchy in a CNI terminal is shown in Figure 3. The control requirements placed on a CNI preprocessor are influenced by the control hierarchy within the terminal and in particular, the control functions that surround the CNI preprocessor in a CNI terminal. At the outer layer, the terminal control function imposes unique requirements to control the levels below it. At the inner layer, the RFG places unique requirements on the CNI preprocessor. Each layer can influence the control architecture of layers above and beneath it. In the following discussion, the terminology "events" and "commands" will be used synonymously. A command will usually cause the invocation of several CNI subfunctions at a prescribed time within a preprocessor. An example is the IFF interrogate receive command, which will invoke the envelope detector, pulse shape detector, pulse position demodulation, SIF filter, and the output interface at a designated time of day. The unique control requirements imposed by the various elements in the CNI terminal on the preprocessor control function will be described next.

System Terminal Controller Requirements

The following requirements are imposed by the terminal controller on the preprocessor control structure.

- Capability to interleave time tagged and immediate commands
- Ability to queue up commands for sequential execution
- Ability to prioritize CNI commands
- Precise start of command execution (nanoseconds resolution)
- Time ordered execution of commands from different prioritized queues without loss of output data
- Detection of stale and invalid commands and queuing errors
- Ability to insert and purge queued commands
- Match commands and associated data when command and data sources are different units within the system
- Ability to initiate command execution when only a part of the command is available while the rest of the command is in the process of being transferred from the terminal control function

Signal Processor Requirements

The signal processor processes CNI preprocessor outputs and also provides transmit data. It performs both encoding and decoding functions, implements various signal processing algorithms, and detects conditions that require immediate terminal response/action (e.g., RTT in a JTIDS system, Mode S response). It places the following demands on the CNI preprocessor:

- Expedient transfer of preprocessor output to minimize system latency and meet turnaround requirements
- Orderly transfer of multiple data message types generated by simultaneously active CNI functions within the preprocessor
- Formatting of output messages in the preprocessor

CNI Subfunction Requirements

The CNI subfunctions (e.g., PN despreading, phase rotation, tracking, and demodulator) within the preprocessor place by far the most demanding requirements on the control architecture. No compromises can be made in this area. Some of the more demanding requirements are listed below.

CNI Subfunction Input Requirements

- Transfer of parameters (e.g., PN for despreading) at precise instants with nanosecond resolution
- Selective parameter updates as a function of events occurring within the CNI subfunction and as a result of SP processing of data (e.g., updates to thresholds for reply rate limiting for IFF transponders)
- Reset capability that includes resetting multiple CNI subfunctions simultaneously, or selected subfunctions as required by the overall CNI sequencing constraints
- Data and command distribution while a CNI subfunction is in progress
- Reconfiguration while a CNI function is active
- Accommodate parameter transfer for the immediate next reconfiguration while mode transition is in progress at the CNI function (e.g., reload correlator with CCSK reference patterns as soon as preamble detection is complete)

CNI Subfunction Output Requirements

The outputs generated by the CNI functions can fall into a variety of types:

- Steady stream of outputs (e.g., AGC samples during GPS)
- Burst outputs (e.g., IFF replies)
- Mixture of above
- All of above with a timeout constraint
- Multiple types of simultaneous outputs from CNI subfunctions

Off-line and On-line Test Requirements

The need for minimizing fault latency and support for a two level maintenance concept in future avionics terminals calls for a variety of fault detection and recovery methods. These place additional demands on the preprocessor control architecture that must now support:

- Immediate detection and reporting of faults
- Orderly generation of on line status reports, including fault status
- Transfer of exception status during CNI function operation
- Off-line functional test capability



Figure 3. Control Hierarchy In A CNI Terminal

Proposed Distributed Control Architecture

Figure 4 shows the proposed control architecture and the allocation of the above control requirements to the architectural elements. Again, a distributed control approach is chosen to implement these various requirements as a centralized controller cannot perform all the control functions satisfactorily while providing controllability and observability. The basic control elements include:

- In-bound Command/Data Router
- Precision Time-Keeping Function
- Input Command Handler
- Output Data Handler
- Out-bound Message Router

The In-Bound Command/Data Router

The In-bound Command/Data Router interfaces to the high-speed buses, sorts incoming, interleaved, prioritized commands received from the buses, and routes the commands and data to the appropriate priority FIFOs. It implements the bus protocol, performs message error detection (loss of words, parity error, missing start, and end of blocks, etc.) and identifies the same for the commands and data queued in the FIFOs. It also flags the start of command for use by the Input Command Handler.

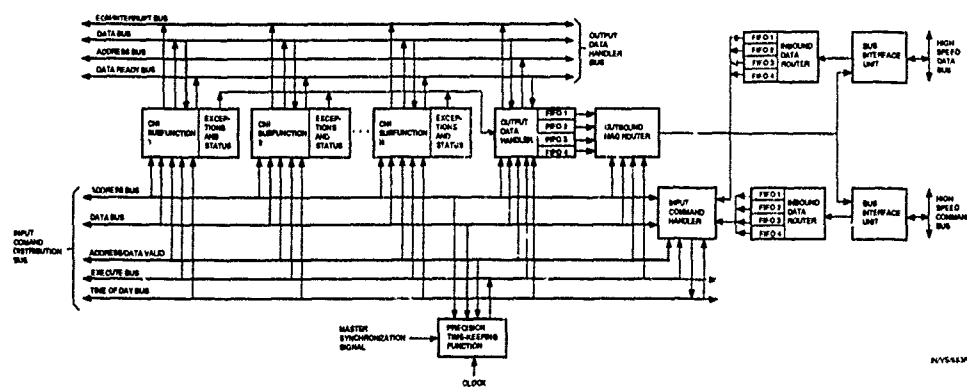


Figure 4. Distributed Control Architecture For The CNI Preprocessor

Precision Time-Keeping Function

The Precision Time Keeping Function provides the capability of matching specified event time of execution with the terminal time of day to effect precise start of CNI event execution via the generation of an execute command pulse to the selected CNI subfunction elements. This function capability to load the specified event time, receive a master synchronization signal to maintain precision time, and provide a precision terminal time (at least several LSBs) for time tagging functions. It also performs windowing around event timetags to determine stale or late commands. It works under control of the Input Command Handler (ICH).

Input Command Handler

The Input Command Handler (ICH) detects invalid/incomplete commands in the queues, sorts commands from queues in time order, executes immediate commands, detects exception conditions (stale, late, too early, etc.), and performs parameter setup as necessary. To implement all of the control requirements listed above, all CNI subfunction elements and their internal parameter stores within the preprocessor are designed as addressable elements. Broadcast as well as multicast addresses are added to the derived address space via the above process. The ICH can thus use a bussing scheme to implement the requirements listed above. A 12 bit address bus, a 16-bit data bus, a data valid control signal, and a reset control signal are used to address any CNI subfunction and the parameter store within each CNI subfunction. The address bus identifies a CNI subfunction element (several MSBs) and the appropriate parameter store within the CNI subfunction element (several LSBs), and the data bus contains the data/parameter. This facility provides random addressing capability to synchronize groups of autonomous functional elements in the architecture.

The ICH can thus sequence the event setup operation, start the operation, update parameters as a function of internal status generated by the CNI subfunction element (e.g., preamble acquisition completion, ready to receive CCKS reference patterns for JTIDS demodulation), resume execution of a suspended command as a result of a new command (e.g., start demodulation command), and stop CNI function operation. Note that the ICH starts setup for each event just prior to the event start time. It uses the precision time-keeping function to compare the event timetag with the current terminal time. It also provides exception status to the output data handler as exception conditions are detected (abrupt end of parameters, incomplete FN, etc.) in the ICH. The ICH also initiates an orderly stop of CNI subfunctions.

Output Data Handler (ODH)

The CNI preprocessor has capability to generate multiple streams of output. To accommodate the various data and data rate requirements, each CNI subfunction element that generates data is required to provide adequate buffering of data and transfer a selected data word on demand by the ODH. The ODH 1) performs message formatting while meeting the throughput requirements of each active CNI function in the preprocessor, 2) uses the concept of a tri-state output data bussing scheme to receive data blocks from each CNI subfunction element, 3) receives a unique data ready flag when a block of data is ready, 4) processes the data ready flags from the various sources using a specified algorithm (first come first served, LIFO, Priority, or based on the mix of active CNI functions etc.) and selects a particular block for transfer, and 5) identifies the block selected and outputs the address of the word in the data block to identify the required data word for formatting. The addressed CNI subfunction element decodes the address and enables the addressed 16-bit data word onto the tri-state data bus. Thus, the ODH selects a data source that has data ready for transfer, sequences the data words of a block, and routes the data to an appropriate FIFO. For fixed block messages, the ODH performs the detection of the last data word in the message and hands off the message to the Out-bound Message Router function. For variable length messages, the data sources can initiate an end of message by asserting an end of message signal. The ODH is designed to handle the worst-case output data generation scenario in the preprocessor. The ODH also provides a means for the CNI subfunction elements to detect overrun conditions. The CNI function elements can declare message overrun if the ODH has not acquired the last word of the data block and a new data block becomes ready to transfer.

Out-bound Message Router

This function gets ready to transfer messages, prioritizes the messages, and appends the control header words prior to initiating transfer to the bus interface unit(s). It breaks messages into segments as required for transfer to meet the system throughput requirements of each function. This function interfaces with two bus interface units and performs simultaneous transfer of different messages to the two bus unit interfaces.

Precision Time-Keeping Function

The Precision Time-Keeping Function provides the capability of matching specified event time of execution with the terminal time of day to effect precise start of CNI event execution via the generation of an execute command pulse to the selected CNI subfunction elements. This function has the capability to load the specified event time, receive a master synchronization signal to maintain precision time, and provide a precision terminal time (at least several LSBs) for time tagging functions. It also performs windowing around event timetags to determine stale or late commands. It works under control of the Input Command Handler (ICH).

Input Command Handler

The Input Command Handler (ICH) detects invalid/incomplete commands in the queues, sorts commands from queues in time order, executes immediate commands, detects exception conditions (stale, late, too early, etc.), and performs parameter setup as necessary. To implement all of the control requirements listed above, all CNI subfunction elements and their internal parameter stores within the preprocessor are designed as addressable elements. Broadcast as well as multicast addresses are added to the derived address space via the above process. The ICH can thus use a bussing scheme to implement the requirements listed above. A 12-bit address bus, a 16-bit data bus, a data valid control signal, and a reset control signal are used to address any CNI subfunction and the parameter store within each CNI subfunction. The address bus identifies a CNI subfunction element (several MSBs) and the appropriate parameter store within the CNI subfunction element (several LSBs), and the data bus contains the data/parameter. This facility provides random addressing capability to synchronize groups of autonomous functional elements in the architecture.

The ICH can thus sequence the event setup operation, start the operation, update parameters as a function of internal status generated by the CNI subfunction element (e.g., preamble acquisition completion, ready to receive CCSK reference patterns for JTIDS demodulation), resume execution of a suspended command as a result of a new command (e.g., start demodulation command), and stop CNI function operation. Note that the ICH starts setup for each event just prior to the event start time. It uses the precision time keeping function to compare the event timetag with the current terminal time. It also provides exception status to the output data handler as exception conditions are detected (abrupt end of parameters, incomplete PN, etc.) in the ICH. The ICH also initiates an orderly stop of CNI subfunctions.

Output Data Handler (ODH)

The CNI preprocessor has capability to generate multiple streams of output. To accommodate the various data and data rate requirements, each CNI subfunction element that generates data is required to provide adequate buffering of data and transfer a selected data word on demand by the ODH. The ODH 1) performs message formatting while meeting the throughput requirements of each active CNI function in the preprocessor, 2) uses the concept of a tri-state output data bussing scheme to receive data blocks from each CNI subfunction element, 3) receives a unique data ready flag when a block of data is ready, 4) processes the data ready flags from the various sources using a specified algorithm (first come first served, LIFO, Priority, or based on the mix of active CNI functions etc.) and selects a particular block for transfer, and 5) identifies the block selected and outputs the address of the word in the data block to identify the required data word for formatting. The addressed CNI subfunction element decodes the address and enables the addressed 16-bit data word onto the tri-state data bus. Thus, the ODH selects a data source that has data ready for transfer, sequences the data words of a block, and routes the data to an appropriate FIFO. For fixed block messages, the ODH performs the detection of the last data word in the message and hands off the message to the Out-bound Message Router function. For variable length messages, the data sources can initiate an end of message by asserting an end of message signal. The ODH is designed to handle the worst case output data generation scenario in the preprocessor. The ODH also provides a means for the CNI subfunction elements to detect overrun conditions. The CNI function elements can declare message overrun if the ODH has not acquired the last word of the data block and a new data block becomes ready to transfer.

Out-bound Message Router

This function gets ready to transfer messages, prioritizes the messages, and appends the control header words prior to initiating transfer to the bus interface unit(s). It breaks messages into segments as required for transfer to meet the system throughput requirements of each function. This function interfaces with two bus interface units and performs simultaneous transfer of different messages to the two bus unit interfaces.

Conclusion

This paper has briefly described a distributed control architecture for an integrated CNI preprocessor. The proposed architecture is simple, effective, and allows multiple CNI functions to co-execute in the same unit while providing flexibility to program, test, and control the unit. It uses the distributed control architecture to process inputs, control autonomously operating CNI subfunctions, and provide outputs to the rest of the terminal.

References

- 1 Subramanyam, V. R., Stine, L. R., Nicklow, V., High Speed Bus for CNI/EW Applications (NAECON 1984)
- 2 Campbell, M. E., Camana, P. C., Subramanyam, V. R., Control Concepts and Architecture for Advanced Multi-role Avionics Terminals (NAECON 1985)
- 3 Subramanyam, V. R., Stine, L. R., Design for Testability for Future Digital Avionics Systems (MILCOM 1985)

JTIDS RELATIVE NAVIGATION - PRINCIPLES, ARCHITECTURE AND INERTIAL MIXING

by
W. R. Fried
Senior Scientist
Hughes Aircraft Company
Fullerton, California
92634
USA

INTRODUCTION. The time-synchronous nature and excellent time-of-arrival (TOA) measurement accuracy of the Joint Tactical Information Distribution System (JTIDS) gives rise to its inherent capability to provide high accuracy relative and absolute navigation information to vehicles carrying JTIDS terminals. This is accomplished through the inclusion of the Relative Navigation (RELNAV) software in the terminal's operational computer program. This function has been called "relative navigation" because it is based on passive and active ranging to the other terminals in the net. If some of these terminals have knowledge of their geodetic position (which may be true for both stationary or moving terminals), the user can determine his absolute position in standard geodetic coordinates, as well as his relative position in an arbitrarily established grid. In order to provide the highest possible accuracy for high dynamic users, the JTIDS RELNAV function is typically interconnected with a dead reckoning system on the vehicle, such as an inertial navigation system. The dead reckoning data is mixed with the JTIDS TOA data in a recursive (e.g., Kalman) filter mechanization to determine the user's position, velocity, and time bias. This paper describes the principles of operation and architecture of the JTIDS RELNAV function and the system configuration for a particular inertial interconnection. The RELNAV observation model, and coordinate systems are discussed. Finally, some simulation results of the system are presented.

THE JOINT TACTICAL INFORMATION DISTRIBUTION SYSTEM (JTIDS). JTIDS is a synchronous, time division multiple access (TDMA), spread spectrum, secure, communication system which also inherently provides high accuracy navigation and identification functions. The system is synchronous since all users in a net operate on a common, precise time base, through use of stable clock oscillators. The clocks are synchronized to a time master called Net Time Reference (NTR) by means of either an active round trip timing (RTT) technique or a passive technique. The latter is an important byproduct of the Relative Navigation function, as will be seen later. The time division multiple access property of JTIDS results from the use of an epoch of time of 12.8 minutes, which is divided into 12-second cycles (frames), and these are subdivided into 7.8125 ms time slots. (See Figure 1.) Several fixed format messages have been defined, including the so-called Precise Position Location Identification message (PPLI-message), which is of particular importance to the Relative Navigation (RELNAV) function, since it contains the position of the transmitting unit. Typically, each unit transmits such a PPLI-message on a periodic basis, each 12-second frame. Detailed descriptions of the JTIDS design characteristics, such as operating frequency band, modulation waveform, etc., are given elsewhere, (e.g., Reference 1) and will not be repeated in this paper.

RELNAV PRINCIPLES. The primary purpose of the JTIDS RELNAV function is to determine own unit position, velocity, and altitude in both relative and absolute coordinates from data received from other terminals, for such applications as vehicle navigation, transmission for command and control, target designation and weapon delivery. The RELNAV function also provides own terminal clock correction for time synchronization. In the JTIDS RELNAV concept, each user passively ranges, sequentially, to several other terminals in the net, by means of a measurement of the time-of-arrival (TOA) of the PPLI-messages transmitted by these terminals, and determines own position from these measured ranges using a form of multilateration called pseudo-ranging. (See Figure 2.) Since JTIDS is a synchronous system, all terminals transmit at specified and known times. Thus, if the user were in perfect synchronization with his sources, three such range measurements (in suitable geometry) could determine his three-dimensional position. However, if the user has a synchronization error (time bias) with respect to the system time, as is normally the case, he needs to determine this time bias, since range and time errors are not separable in a single passive ranging event (TOA measurement). A sequence of such passive range measurements on signals from sources having higher position and time quality will provide a continuously updated measure of the user's position and time bias. (This time bias determination is also called passive synchronization.) The total process is called pseudo-ranging, since the range measurements are not "true" ranges, but are made with respect to the user's own clock time, rather than on an absolute basis.

The basic RELNAV observation model can be expressed by

$$R_0 = c (TOA) = R_c + b_s - b + N \quad (1)$$

and

$$R_c = \left[(X_s - X)^2 + (Y_s - Y)^2 + (Z_s - Z)^2 \right]^{1/2} \quad (2)$$

where: R_0 is the observed range (pseudo-range) to the source; c is the speed of light; TOA is the observed time-of-arrival with respect to the user's own clock time, R_c is the computed (predicted) range to the source, based on the best estimate of own position and the source's transmitted position; X_s, Y_s, Z_s are the source's transmitted position coordinates; X, Y, Z are the user's own position coordinates; b_s is the residual time bias of the source; b is the time bias of the user; N is the sum of measurement noises. Each unit attempts to zero his time bias before transmission. Thus, the residual source time bias b_s is very small, is not known to the source and cannot normally be estimated by the user. Hence, it is considered included in the measurement noise N and the basic RELNAV observation model becomes

$$R_0 = c(\text{TOA}) = R_c - b + N$$

(3)

The basic observable is the TOA measurement. With the source positions known from the data in the received PPLI-messages, several measured pseudo ranges R_0 (equation 3) will lead to a determination of the user's own position coordinates X, Y, Z , and his time bias b . Since the data is obtained sequentially, recursive filtering is used in the RELNAV mechanization and for most applications, dead reckoning (e.g., inertial) data is mixed with the TOA data for smoothing and prediction purposes (see Figure 3). A RELNAV Kalman filter can then inherently estimate the dead reckoner (inertial) errors using the accurate TOA measurements as reference data, thereby continuously "calibrating out" these dead reckoner errors. It is also possible to operate RELNAV using TOA derived data only (see Reference 2), wherein position and velocity, and possibly also acceleration, may be carried as filter states. This form of RELNAV operation is more applicable to relatively low-dynamic vehicles such as ships and ground vehicles, or low-dynamic aircraft. In high-dynamic vehicles this mode of operation would lead to excessive lags in position determination during maneuvers. The user terminal receives the reported source position, and time and position qualities in the received PPLI-messages from other terminals. The RELNAV Source Selection function (see Figure 4) then selects the desired sources, based on the relative qualities, and geometric considerations with respect to the terminal's own error ellipse. The terminal measures the TOAs of the received messages and the RELNAV function uses those of the selected sources. The RELNAV filter then calculates predicted TOAs from extrapolated position data and from the received source position data. It compares the measured and predicted TOAs to generate residuals. Using each of these observations, the RELNAV filter updates the various states of the filter, such as position, velocity and time errors as well as the covariance matrix. After several such observations have been used, the filter applies corrections to the Navigation Data Extrapolation function, which continues to extrapolate the navigation data with these latest corrections, for use on transmitted PPLI-messages, vehicle avionics and for subsequent filter prediction purposes. The control data from the filter representing clock corrections (i.e., time bias and frequency errors, are provided to the terminal clock. In the filtering process, the reported time and position qualities of the sources are converted to equivalent variances and are used to modify the appropriate terms in the filter gain calculation for each observation. Also, the non-linear observation equation (equation 3) is linearized to calculate the filter gains.

In addition to the basic passive ranging operation just described, the RELNAV function also uses an active or round trip timing (RTT) technique available in JTIDS terminals, provided the unit is designated as an active (non-radio silent) terminal. This technique independently generates an excellent measure of the user's time bias. In the RTT technique, the user transmits a message to a selected unit of higher time quality (called the donor); the donor measures the time-of-arrival (TOA) of that message and reports it back to the interrogating user unit in a return message. The latter then measures the TOA of the donor's reply message. The difference between the reported TOA and the user measured TOA (divided by 2 and corrected for a known fixed transponding delay) is a very accurate measure of the time offset of the user with respect to the clock time of the donor. If the donor is the net time reference (NTR), the measured time offset is the desired time bias b of equation 3. Thus, to optimize RELNAV system operation, it is intended to permit some units to RTT relatively frequently (i.e., whenever their filter time variance calls for it), in order to maintain the highest time quality, so that they, in turn, can be used as "navigating sources" by other units, which may have to be radio-silent for tactical reasons.

RELNAV ARCHITECTURE AND COORDINATE SYSTEMS. The JTIDS RELNAV function operates in two coordinate systems, i.e., in the conventional geodetic (latitude-longitude) system, and in an arbitrarily established "relative grid" system. It can operate either simultaneously in both systems or in either system alone. The RELNAV "relative grid" is defined on the basis of a Cartesian U/V coordinate frame, with the UV plane tangent to the geoid at the true grid origin and the V coordinate nominally (though not exactly) being in the direction of the meridian (True North) at the grid origin (see Figure 5). The manner of establishing the grid origin and grid north direction is discussed later in this section. A hierarchy of terminal categories has been established to permit operation in these two coordinate systems in an optimum manner. In either coordinate frame, one member of the net is designated as the NTR, wherein his clock establishes system time. The NTR is assigned the highest Time Quality, i.e., 15 from a possible range of 0 to 15. Terminals which possess independent, accurate knowledge (i.e., better than 50 ft 1 σ) of their geodetic position, such as surveyed ground stations, are designated as Geodetic Position References (GPR) and are assigned the highest Geodetic Position Quality (Q_{gp}) i.e., 15 from the range of 0 to 15. In relative grid operation, the grid origin and grid north direction are established either by a single moving terminal, designated as the Navigation Controller (NC) or by two stationary terminals designated as NCs. An NC is assigned the highest Relative Position Quality (Q_{pr}), i.e., 15 from a range of 0 to 15, and the highest Relative Azimuth Quality (Q_{ar}), i.e., 7 from a range of 0 to 7. Operationally, the one moving NC case is probably the more typical one. The NC's transmitted relative position coordinates U and V are perfect (error-free) by definition. He defines the grid origin by virtue of his reported U and V coordinates, his actual geodetic position and his dead reckoning (e.g., inertial) navigation system's estimated north direction. The grid V-direction is intended to be generally north-pointing. Initially, the single moving NC establishes the grid origin either where he believes to be located (local grid origin) or at some arbitrary offset point (offset grid origin). Thereafter, the NC's dead reckoning navigation system's velocity errors will give rise to a lateral motion of the grid origin. Similarly, the NC's dead reckoner's heading error will give rise to a grid offset angle β which is non-zero. Unless the NC has accurate geodetic observations available to him, he himself will not know the true value of β , nor will other terminals which are ranging to him without having accurate geodetic information. However, as a result of their sequential ranging to the NC (and to other terminals in the relative grid) these terminals will achieve accurate relative navigation within the established common grid, without the availability of geodetic references (e.g., over the ocean). However, whenever access to geodetic position references is available to a terminal, as in the case of over-land missions, RELNAV will also provide high accuracy absolute geodetic navigation. The two stationary NC case operates somewhat differently from the single moving NC case. Here, each NC is still assumed to report perfect (error-free) U and V coordinates. These four elements of data establish the three grid parameters, i.e., the two grid origin coordinates and the grid offset angle, β . In order to achieve internal consistency, a fourth parameter must be independently known, namely, the true range between the two NCs. In JTIDS, this can be achieved by means of TOA measurements.

Below the NTR, GPRs, and NCs, there are several classes of users in the RELNAV architecture which differ on the basis of reported time and position qualities and the manner of time synchronizations which they are permitted to use. They are called Primary User (PU), Secondary User Active (SUA), and Secondary User Passive (SUP). Only the NTR has a Q_t of 15; only the GPRs have a Q_{pg} of 15 and only the NCs have a Q_{pr} of 15. GPRs, NCs and PUs are permitted to RTT whenever the filter time variance indicates that they have a time quality one half level below the best available source. Thus, the quality of these units is maintained as high as possible. The SUP terminals are not permitted to RTT at all, i.e., they are required to perform navigation and synchronization only passively. SUA type terminals are intended to perform navigation and synchronization primarily using passive operation, but are permitted to RTT infrequently, on a periodic basis. Regardless of the class of the general user, the reported quality levels are established on the basis of the respective variances calculated by the user's RELNAV Kalman filter. The variance-to-quality level conversion algorithms have been standardized. Certain minimum source selection constraints have also been established which are designed to assure stability of operation by preventing units from using sources for position and time updates, which have lower position and time qualities than they do themselves. Without such constraints, it is possible to have divergence of time and position errors (Reference 4). If a user has been able to determine his position in both the geodetic and relative coordinate systems, he is required to report in his PPLI-message both his geodetic and relative coordinates, his estimate of the beta angle, as well as his Q_{pg} , Q_{pr} , and Q_{ar} . As a result, it becomes possible for another unit to use the combined geodetic and relative position data in that terminal's PPLI-message to initially acquire the grid origin or to improve its knowledge of geodetic and relative position after grid acquisition.

JTIDS RELNAV ERROR CHARACTERISTICS. The achievable accuracy performance of JTIDS RELNAV is a function of several error sources. Table I lists these error sources and shows their characteristics and their impact on RELNAV system performances. The table also indicates the design approaches which have been used in JTIDS terminals and RELNAV implementations, in an effort to minimize the effects of these error sources.

The TOA measurement accuracy represents an ultimate limitation on achievable RELNAV system accuracy. It is a function of the signal bandwidth, the available signal-to-noise ratio and the performance of the TOA measurement circuit. In JTIDS terminals, a high performance delay lock loop is used for this function.

The short term stability of the terminal's clock is of importance for RELNAV performance and its impact increases as received PPLI-message rates are decreased. Table I shows the short term stability of typical clock oscillators used in JTIDS terminals.

TABLE I. RELNAV ERROR SOURCES

ERROR SOURCE	CHARACTERISTICS	PERFORMANCE IMPACT	DESIGN APPROACH
TOA Measurement Error	Hardware and S/N dependent	Ultimate limitation on accuracy	High performance delay lock loop
Short Term Clock Stability	Hardware dependent	Greatest impact at PPLI-message rates	10^{-10} to 10^{-9} for short average times
Position and Time Quality of Sources	Scenario and Net Management dependent	Large impact on user accuracy	Source selection logic based on qualities
Geometric Dilution of Precision (GDOP)	Relative geometry user and sources; scenario dependent	Large impact on user accuracy	Source selection logic based on geometry
Propagation Effects	Caused by atmospheric refraction	Greatest effect for large ranges and atmosphere variations	Use of propagation correction model
Dead Reckoning Sensor Performance	Sensor Dependent; INS most accurate	Greatest effect for high dynamic user and low received PPLI-message rates	User platform dependent
Computational Errors	Processor Dependent	Limits ultimate accuracy	Processor with floating point arithmetic
PPLI-Message Rate	Net Management and scenario dependent	Low rate makes accuracy more dependent on vehicle dynamics; dead reckoner and clock performance	Complete dead reckoner and high performance

The position accuracy achievable by a user at any particular time is clearly a function of the position and time qualities of sources which are in line-of-sight of the user at that time. Thus, there is an inherent dependence on the mission scenarios and the deployment of reference terminals, such as the Navigation Controller and Geodetic Position References. In addition, it is important to select for RELNAV processing the best available sources, from a position and time quality viewpoint. Therefore, the RELNAV algorithms includes a source selection logic which screens the sources based on several categories, i.e., geodetic and relative position, altitude, time and grid azimuth and as a function of relative qualities. The relative geometry between the user and the sources gives rise to the well-known geometric dilution of precision (GDOP) effect, which determines the achievable position accuracy. Conceptually, it can be regarded as a multiplier on the basic ranging error between the user and the sources. One of the major purposes of the source selection function is to select the sources for processing in pairs, so as to minimize this GDOP effect, based on the relative geometry and also on the user's own error ellipse.

The atmospheric density gives rise to an error in range measurement due to the bending and retardation in the radio path. The impact on range error increases directly with range between the source and user. Atmospheric variations can cause significant error deviations from simple atmospheric models. RELNAV algorithms can include a model for the atmospheric radio path variation, which calculates an atmospheric delay value which is then applied to the measured TOA, in order to eliminate a major portion of this error.

If a dead reckoning sensor is interfaced with RELNAV, its errors and their correlation times will affect the resulting position error. However, the impact of the dead reckoner errors is only significant for high-dynamic operation and for low received PPLI-message rates. Modern inertial navigators tend to have better short term accuracy than Doppler navigators and airspeed dead reckoners. For a given application, the choice of dead reckoner for RELNAV interconnection typically depends on the equipment already existing on the platform which is, in turn, a function of the platform's mission and dynamics.

The processing precision of the computer used for RELNAV processing represents an ultimate limit on the achievable RELNAV accuracy performance, given "perfect" sensor performance, i.e., TOA measurement accuracy, clock stability, etc. In current processors the computational error can be expected to be very small compared to the other RELNAV error sources mentioned.

Finally, the higher the PPLI message transmission rates of the members of a RELNAV community, the higher is the potential RELNAV accuracy performance throughout the net. For a particular net, the PPLI-message reception rate for available (and usable) sources is the important parameter. A low received PPLI-message rate makes instantaneous RELNAV position accuracy more dependent on vehicle dynamics, dead reckoner performance and clock stability. The latter is of greater significance for pure passive (radio-silent) operation, since the RTT technique will serve to independently provide clock correction data. Also, a high short-term stability of the terminal clock inherently reduces the effects of low received PPLI-message rates on the user's knowledge of time, and hence also on his ability to determine his position accurately.

RELNAV OPERATIONAL BENEFITS. The architecture described in the previous section leads to a number of unique operational benefits of the JTIDS RELNAV function. RELNAV operates simultaneously in both world-wide-geodetic and arbitrary-grid coordinates. While some units may not have access to accurate absolute geodetic information, they will be well-registered with each other in the arbitrary grid (established by the NC), and possess high relative position and azimuth accuracy within that grid. Thus, in a typical military application, one member of the net may acquire a target in grid coordinates, transmit the target data to another unit, and that unit should be able to attack that target with accuracy. The high relative heading and position accuracy obtained from the RELNAV function is equivalent to a high relative pointing accuracy between users, which is of importance for a number of applications. Pointing is the ability to physically point to another unit, while relative heading error is the error in transferring a bearing measurement from one unit to another. Some military targets are more typically designated in geodetic coordinates, and in this case, the geodetic information available from RELNAV would be used for transfer of target data. RELNAV, of course, inherently exhibits the same high jam-resistant properties as JTIDS per se. Since RELNAV can operate without any fixed ground stations or any fixed-orbit satellites, it is not dependent on any vulnerable nodes, as is the case for other similar TOA measurement based systems. If some units, perhaps near the perimeter of the net, have access to accurate ground-based references, RELNAV operation inherently propagates their improved knowledge of geodetic position throughout the net, well beyond line-of-sight of the ground sites or the units in the vicinity of these sites. Similarly, knowledge of grid position and time can be propagated beyond the horizon, in spite of the high radio frequencies used in JTIDS.

In an inertially-aided RELNAV configuration, if radio data is temporarily lost for any reason or the user leaves the net permanently, the inertial errors will have been accurately estimated (calibrated out), resulting in a much lower error buildup than would be experienced with an inertial navigation system operating by itself. Similar calibration of dead reckoner errors can be obtained when RELNAV is interfaced with such dead reckoning systems as a Doppler radar, airspeed sensor, EM (Electromagnetic) log, and a heading reference. This is achieved by modeling the significant errors of these dead reckoners in the RELNAV filter, estimating them by using the TOA measurements as a reference and applying them as controls (resets) to the navigation data extrapolation function. For example, for Doppler radars, the Doppler scale factor error would be modeled, while for airspeed and EM log sensors the respective scale factor errors would be modeled, along with the heading error.

It is interesting to note that the characteristics of JTIDS RELNAV and the NAVSTAR Global Positioning System (GPS) complement each other in a number of ways (Reference 6). Both are synchronous, spread spectrum systems, using the passive pseudo-ranging technique for position determination. GPS is a global system, while JTIDS RELNAV is a tactical system which provides a local common grid capability. Since both systems are time referenced systems, it should not be difficult to use a common time base, e.g., by synchronizing JTIDS time masters (NTRs) and other JTIDS units to GPS system time via a time strobe and data transfer. Whenever GPS position data is available in one or more units of a JTIDS net, these units can act as Geodetic Position References for RELNAV, as well as providing an excellent time reference. As long as both systems are operational on a particular vehicle, the position and velocity data could be fed from GPS to RELNAV. Inertial aiding could be applied to both systems, as required. If GPS data were temporarily lost (due to jamming, loss of satellites or other reasons), JTIDS RELNAV could provide continuous accurate navigation data and time. It may also be useful to relay GPS satellite ephemeris data over the JTIDS data link. RELNAV time and position data would shorten GPS re-acquisition time (time-to-first-fix) when GPS satellite data again becomes available. Similarly, a priori knowledge of common GPS/JTIDS time could shorten JTIDS net entry and initial synchronization. In some platform implementations, it may be desirable to perform the RELNAV-inertial mixing or the RELNAV-GPS-inertial mixing in a central computer. In that case, a RELNAV algorithm can be envisioned which provides TOA-only derived position and velocity data to a filter in the central computer which then uses these data to obtain optimum estimates of position and velocity and inertial errors. A fully integrated RELNAV-GPS-inertial filter mechanization has been investigated (Reference 5). A combined JTIDS-RELNAV-GPS-Inertial mechanization could provide optimum operation as regards accuracy, coverage and jam resistance for both global and tactical applications. However, a pure RELNAV-Inertial mechanization alone, as described in what follows, can provide high accuracy position, velocity and time within a JTIDS net in, both, relative grid and geodetic coordinates, provided at least two units having good knowledge of their geodetic position are available.

RELNAV-INERTIAL CONFIGURATION. A particular RELNAV-Inertial mechanization was designed at Hughes Aircraft Company for incorporation into the operational computer program (OCP) of the Class 1 HIT JTIDS Terminal on a development program supported by the U.S. Air Force and is described below. RELNAV-Inertial mechanizations have also been designed for other JTIDS terminals (e.g., the Class 2 Terminal) and use somewhat different mechanizations for RELNAV-Inertial mixing, state vector composition and source selection. The fundamental principles are the same, however. The mechanization described herein was designed and tested for interconnection with a particular Inertial Navigation Set (INS); namely the AN/ASN-109, and a particular Air Data Computer (ADC), namely the AN/ASK-6. The latter is required to provide barometric altitude data. In the selected interface approach, north, east and vertical velocity components are received from the INS and fed to the Inertial Navigation Processing Function. Barometric altitude is fed from the ADC to the INS for use in its baro-inertial loop. The functional software flow of the RELNAV-Inertial configuration is shown in Figure 6. JTIDS PPLI-messages are received from the terminal's Message and Signal Processing software function and screened for acceptability in the Source Selection function. Source data and associated TOAs from selected PPLI-messages are then fed to the RELNAV filter for background processing. The Filter receives extrapolated position data (NAV STATES), which are valid at the times of interest, from the Inertial Navigation Processing Function and, in turn, provides the filtered updates (CONTROL STATES) to the Inertial Navigation Processing Function. The outputs include position (geodetic and relative), ground speed, altitude, course, beta angle and the various quality levels (position, time and relative azimuth). These data are computed and formatted for the outgoing PPLI-message, whenever the terminal transmission assignments calls for a PPLI-message transmission. Since the Inertial Navigation Processing Function typically computes the navigation data several times per second, these data are suitable for use by an aircraft avionics system for vehicle navigation and flight control.

STATE VECTOR SELECTION. The fundamental relationship of the RELNAV dual grid navigation problem is

$$\underline{R} = \underline{G} - \underline{O} \quad (4)$$

where

\underline{R} = Relative position vector
 \underline{G} = Geodetic position vector
 \underline{O} = Origin position vector

All three vectors in Equation (4) must be expressed in a common frame. Because both the origin and the geodetic position vector are maintained in geodetic coordinates it was considered algorithmically simpler to estimate errors in these states rather than errors in the relative position state. This selection appeared to simplify computations for all net members except the NC. Because the NC does not perform relative navigation, the increased computational burden could be more easily absorbed by the NC. Given this fundamental choice (Equation 4), following are the error states of the Kalman filter selected for the Hughes Aircraft Company Class 1 HIT JTIDS Terminal RELNAV mechanization:

$\Delta \underline{P}_G$ = Geodetic Position States (3 dimensions)
 $\Delta \underline{V}_G$ = Geodetic Velocity States (2 dimensions)
 $\Delta \underline{\psi}$ = Platform Pointing States (3 dimensions)
 $\Delta \underline{\epsilon}$ = Gyro Bias States (3 dimensions)
 $\Delta \underline{P}_O$ = Origin Position States (2 dimensions)
 $\Delta \underline{V}_O$ = Origin Velocity States (2 dimensions)
 $\Delta \underline{R}$ = Relative Grid Azimuth (scalar)
 $\Delta \underline{C}$ = Clock States (offset and frequency)

These 18 states are based on the geometry of the problem. There may be sufficient community information to preclude the inclusion of the gyro states in the state vector during normal operation. However, their inclusion leads to a better autonomous navigation capability when a user temporarily has no access to JTIDS messages or leaves the net permanently.

SOURCE SELECTION. The source selection algorithm is intended to screen incoming navigation messages and select the best available set of data for inclusion in the filter state update. The particular source selection algorithm developed for the Hughes Aircraft Company Class 1 HIT JTIDS Terminal RELNAV function is described in what follows. Somewhat different source selection algorithms have been designed for other JTIDS terminals, e.g., the Class 2 Terminal. There are several different types of source selection categories for which a message may be selected. These include geodetic level position (2 dimensions), grid level position (2 dimensions), vertical (1 dimension), time (1 dimension) and grid azimuth (1 dimension). Incoming messages are screened against these categories. There are four fundamental filter update types that can be constructed from incoming PPLI-messages as follows.

Grid Pseudo-range - A predicted TOA computed from the source grid position and the internal estimate of own grid position and time.

Geodetic Pseudo-range - A predicted TOA computed from the source geodetic position and the internal estimate of own geodetic position and time.

Grid/Geodetic Offset - A predicted distance error computed from the source geodetic position transformed through the internal estimate of own-ship grid/geodetic transformation and the source grid position. The predicted distance error is then projected along and across the line-of-sight between own-ship and source. This yields two components of distance error. The pseudo-range measurement TOAs are range errors which depend on message TOA. The calculated offset errors are independent of TOA. The filter update types can then be used for the various source selection categories. (See Table II.) In addition, a RTT event can be used to update the time parameters. A source is accepted if its use will improve a source selection category.

TABLE II. FILTER UPDATE TYPES AND SOURCE SELECTION

FILTER UPDATE TYPES				
	<u>Grid Pseudo Range</u>	<u>Geodetic Pseudo Range</u>	<u>Offset Along LOS</u>	<u>Offset Across LOS</u>
<u>SOURCE SELECTION CATEGORIES</u>				
Grid Level Position	X		X	
Geodetic Level Position		X	X	
Vertical Position	X	X		
Time	X	X		
Grid Azimuth				X

Because of geometric reasons the level position criteria are treated as paired events. The selection algorithm is initialized as if there were two acceptable sources (pseudo-sources) along the major and minor axis of the system-estimated error ellipse. Each new source is tested to determine if the error ellipse will be improved by its inclusion. If it is, the source replaces one pseudo-source and all subsequent messages are tested against the new error ellipse. The pseudo-sources are never used by the filter.

SIMULATION RESULTS. Simulation results are shown in Figures 7 through 12 of the particular RELNAV-Inertial mechanization developed at Hughes Aircraft Company for the Class 1 JTIDS HIT Terminal on a development program supported by the U.S. Air Force. Figure 7 depicts a case of geodetic navigation of a Primary User in the absence of a relative grid. The user is traveling east and is in view of two Geodetic Position References (GPRs), located in good geometry with respect to the user. All types of error sources are included in the user's system. The initial user position errors are 2350 feet in the east direction and 914 feet in the north direction. The sources transmit PPLI-messages at a 12-second rate. It is seen from Figure 7 that the user's north and east position errors converge to less than 100 feet in less than 2 minutes. Figure 8 depicts a case of relative grid navigation in the absence of geodetic sources, except during the grid acquisition period, when access to both geodetic and relative sources is a basic requirement. For purposes of grid acquisition, during the first minute of the mission, both of the sources are assumed to have good knowledge of their geodetic and relative position and report these in their PPLI-messages. As soon as the user had acquired the grid (in less than one minute), the simulation artificially cut off the geodetic data from the PPLI-messages of the sources. The results in Figure 8 show that the user maintains good relative grid accuracy, i.e., his U- and V-errors are less than 100 feet, while his geodetic errors (only the north error is shown) continue to increase as a result of the normal position error buildup of his inertial system. During this period, the user's errors in geodetic position and grid origin position are highly correlated, keeping his relative grid errors small, as intended for grid RELNAV operation.

Figures 9 through 12 show some simulation results for scenarios where the user transitions from pure geodetic to dual grid operation and from pure relative to dual grid operation, as well as dual grid and pure relative scenarios. For example, from Figure 9 it is seen that the geodetic errors converge to very small values after approximately 2.5 minutes of operation. Similarly, the relative errors converge shortly after the Navigation Controller has entered the net. In Figure 10, the reverse was tested, starting out and converging in pure grid operation, with all terminals moving, and with a Geodetic Position Reference (GPR) entering the net after 6.3 minutes. Again, stable performance is exhibited throughout the simulated mission showing high position accuracies (near 100 feet). These simulation scenarios actually were not particularly optimum for geodetic navigation, since only one GPR was present in the net; however, the motion of the users served to bring down the errors. Figures 11 and 12 show scenarios for pure relative grid operation and dual grid operation, using different user-to-source geometries. These results were obtained from non-real time simulation. The system was also tested with real time software and an actual Class 1 JTIDS HIT Terminal and inertial equipment hardware, showing good performance results.

CONCLUSIONS

JTIDS-RENAV not only accurately determines each unit's position and velocity in an onboard process, but also transmits these data on a secure link to other members of the community of users. As a result, JTIDS RELNAV should find application not only for navigation, target data transfer for weapon delivery, identification and command and control, but also for such related uses as air-derived and ground-derived collision avoidance, air traffic control, rendezvous, station keeping and missile guidance.

REFERENCES

1. Dell-Imagine, R. A., "JTIDS - An Overview of the System Design and Implementation" Proceedings of the IEEE Position Location and Navigation Symposium, San Diego, CA; IEEE Publication 76 - CH1138-7 AES, pp. 212-218, November 1976.
2. Fried, W. R., "Principles and Simulation of JTIDS Relative Navigation", IEEE Transactions, Vol. AES-14, No. 1, January 1978.
3. Fried, W. R. and Loeliger, R., "Principles System Configuration and Algorithm Design of the Inertially Aided JTIDS Relative Navigation Function", Navigation, Journal of the Institute of Navigation, Vol. 26, No. 3, Fall 1979.
4. Westbrook, E. A. and Snodgrass, R. C., "Relative Navigation by Passive Ranging in a Synchronous Time Division Multiple Access Data Net", The MITRE Corporation MTR-2996, prepared for Electronic Systems Division, U.S. Air Force, March 1975.
5. Kriegsman, B. A. and Stonestreet, W. M., "A Navigation Filter for an Integrated GPS-JTIDS-INS System for Tactical Aircraft", Proceedings of the 1978 IEEE Position Location and Navigation Symposium, San Diego, CA, November 1978.
6. Fried, W. R., "Operational Benefits and Design Approaches for Combining JTIDS and GPS Navigation", Navigation, Journal of the Institute of Navigation, Vol. 31, No. 2, Summer 1984.

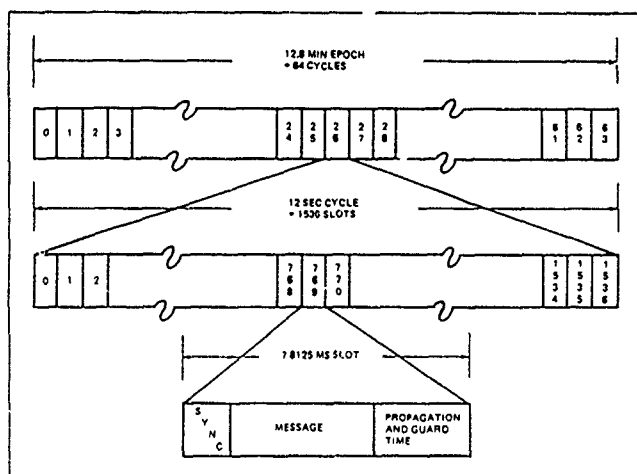


Figure 1. JTIDS Message Timing

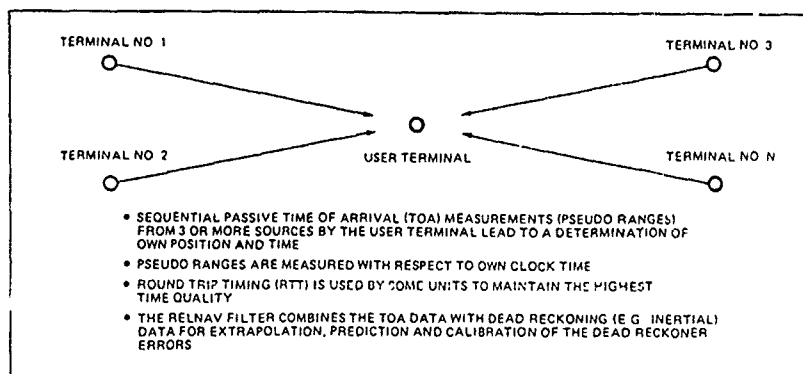


Figure 2. Basic JTIDS RELNAV Principles

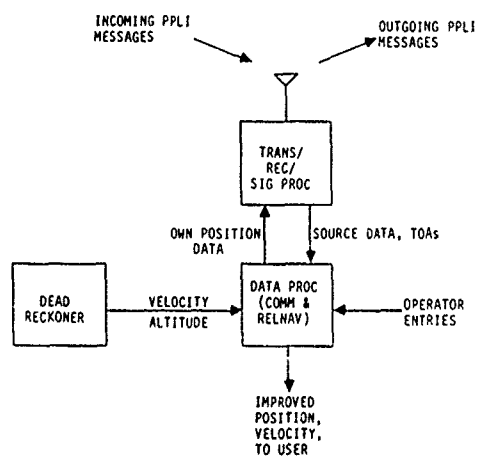


Figure 3. JTIDS RELNAV Operation

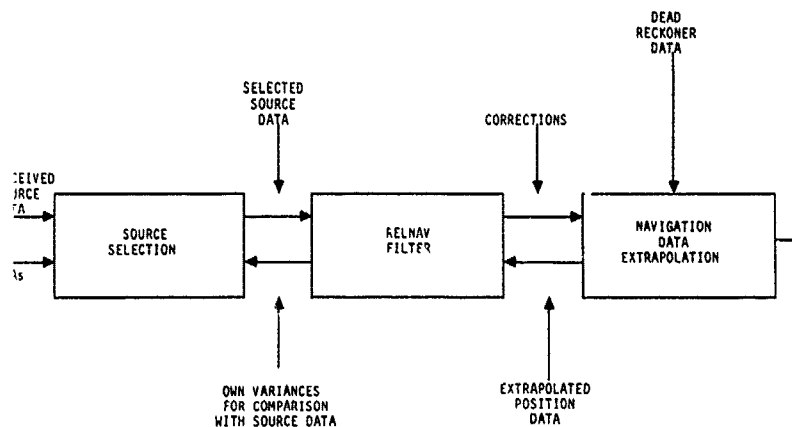


Figure 4. Major RELNAV Processing Function

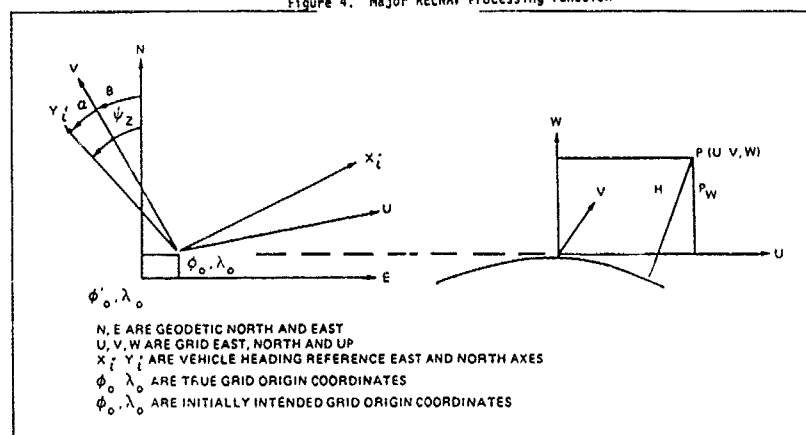


Figure 5 Relative Grid-to-Geodetic Coordinate Relationships

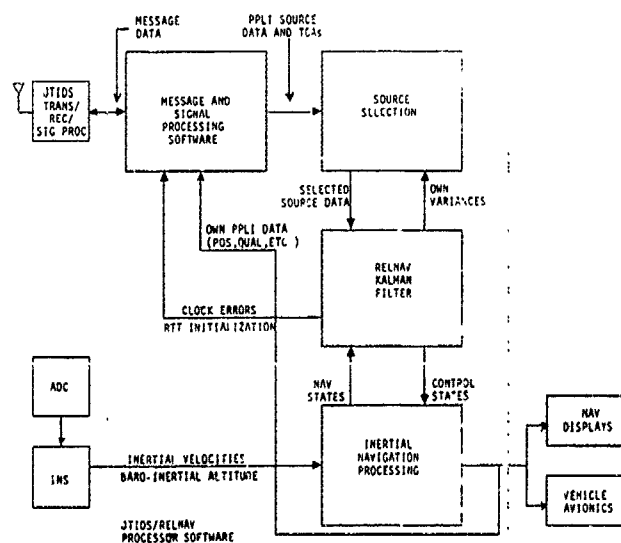


Figure 6 RELNAV-INERTIAL Functional Software Flow

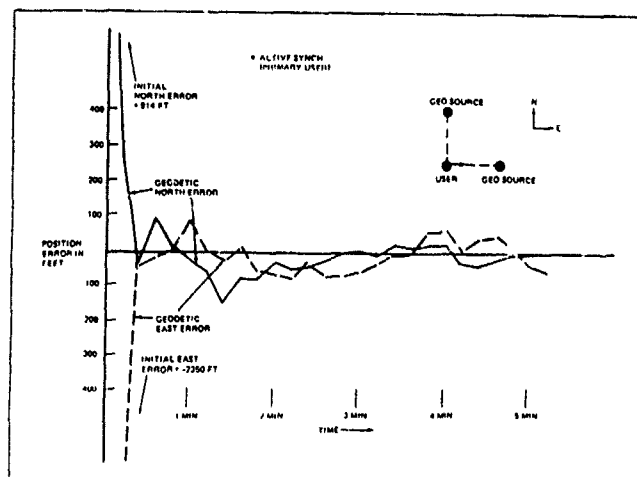


Figure 7 Geodetic Navigation in the Absence of a Relative Grid

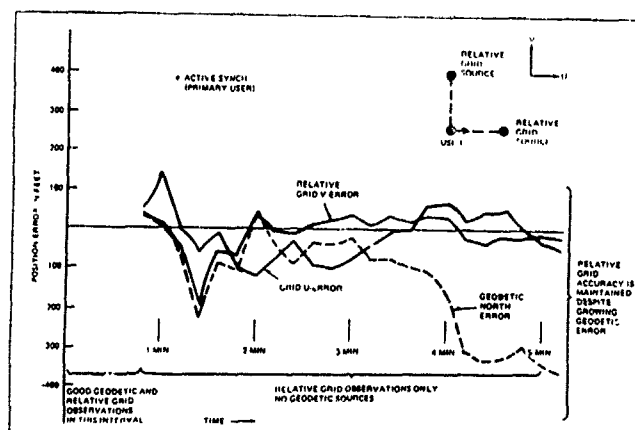


Figure 8 Relative Grid Navigation in the Absence of Geodetic Position Sources

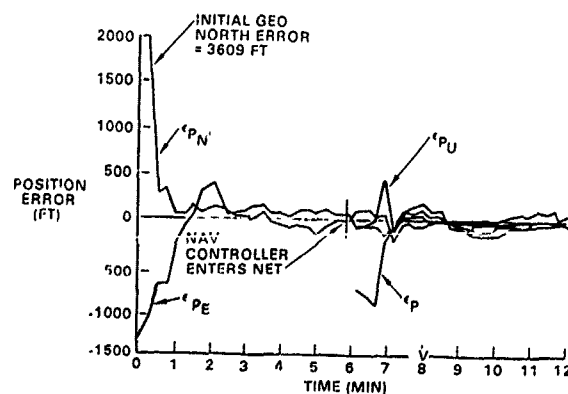


Figure 9 Changing Pure Geo-to-Dual Grid Scenario Simulation Results

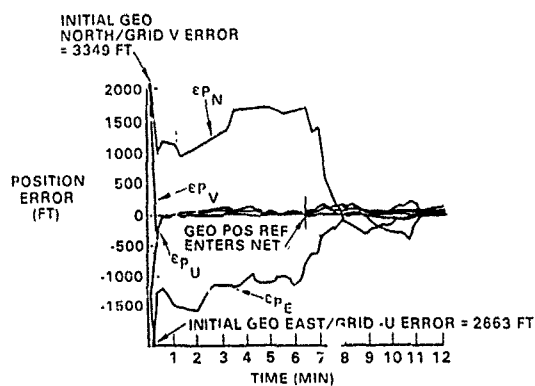


Figure 10 Changing Pure Relative-to-Dual Grid Scenario Simulation Results

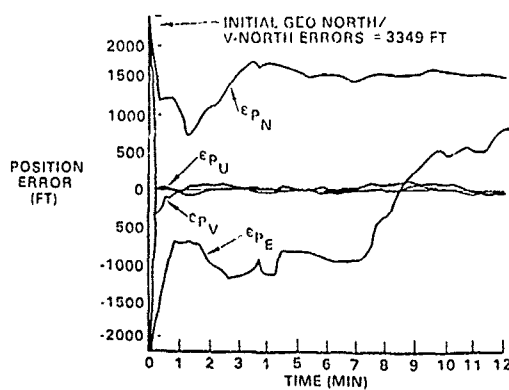


Figure 11 Pure Relative Grid Operation Simulation Results

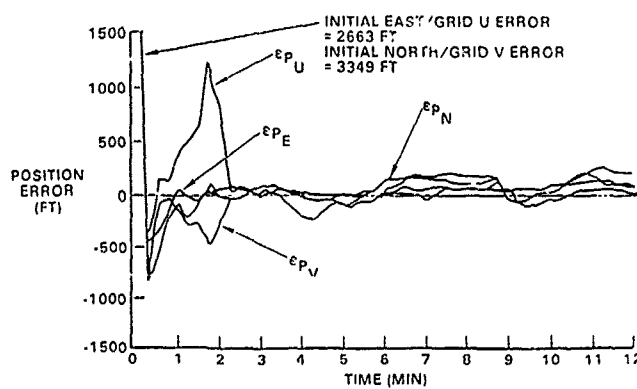


Figure 12 Dual-Grid RELNAV Operation Simulation Results

PLRS — A NEW SPREAD SPECTRUM POSITION LOCATION REPORTING SYSTEM*

by

James A. Kivett and U.S. Okawa
Hughes Aircraft Company
Communications Systems Division
Fullerton, California 92634
United States

ABSTRACT

The Position Location Reporting System (PLRS) provides position location, tracking, and reporting for communities of hundreds of cooperating users in a tactical environment. PLRS uses time-of-arrival (TOA) measurements aided by barometric pressure measurements to establish position tracking of these large communities of user terminals. The system uses the positions of a few user terminals as grid references so that all positions are available both to the cooperating users and to command centers. All control, measurement reporting, and data exchange are cryptographically secured in a synchronous anti-jam communications network.

INTRODUCTION

PLRS will provide the Army and Marine Corps with a unique range of capabilities which will, in many respects, alter the conduct of battlefield operations. For the individual tactical user, either on foot, in a surface vehicle, or airborne, the system determines and displays to him his accurate position location in real-time. It alerts him if he enters a restricted area. It also provides him with guidance to predesignated points, to other units, or along corridors in accordance with his requests, as well as providing a limited free text data exchange capability.

For the tactical commander, PLRS provides the identification, location, and movement of all cooperating users within an assigned area of responsibility. In addition to allowing the commander to monitor the movement of his forces, PLRS provides him with the ability to input and modify coordination points, corridors, and restricted zones.

For all participants, the system (which operates beyond line-of-sight via integral relays) incorporates effective electronic counter-counter measures (ECCM) and provides cryptographically secure digital data communications. Each user has the capability of sending preassigned 2-character messages to provide data to or request information from the system, as well as 12 character free text message exchanges.

PLRS operates both ashore and afloat under all conditions of visibility, weather, and terrain. Its configuration ensures continuity of operation during the transitions of tactical headquarters (e.g., command post displacements, ship-to-shore movement, and passage of lines) and allows for survivability even if a major system element becomes inoperative.

The single community can support an Army Division or Marine Amphibious Brigade with a variable distribution of manpack, surface vehicle, and airborne users. Full system performance is provided within a 47-km by 47-km primary operating area, and airborne users can be located and tracked (with slightly reduced accuracy) within a 300-km by 300-km extended operating area. It can inter-operate with at least four other PLRS communities in adjacent geographical areas.

PLRS incorporates provisions for: quickly accommodating transient users entering the system's tactical area of responsibility; allowing two or more PLRS communities to maintain positioning and identification information for a single user or group of users; and providing automatic or manual transfer of a user from one PLRS community to another.

MAJOR ELEMENTS OF PLRS

PLRS employs two categories of hardware, Master Stations (MS) and user units (UU) to provide the capabilities described above. The two forms of hardware operate in a synchronous time division multiple access (TDMA) network structure. Figure 1 illustrates the mix of equipment that makes up a PLRS network.

*This work was supported by the U.S. Army and U.S. Marine Corps under contract DAAB-07-76-C-1750 administered by the U.S. Army Communications, R & D Command, Ft. Monmouth, N.J.

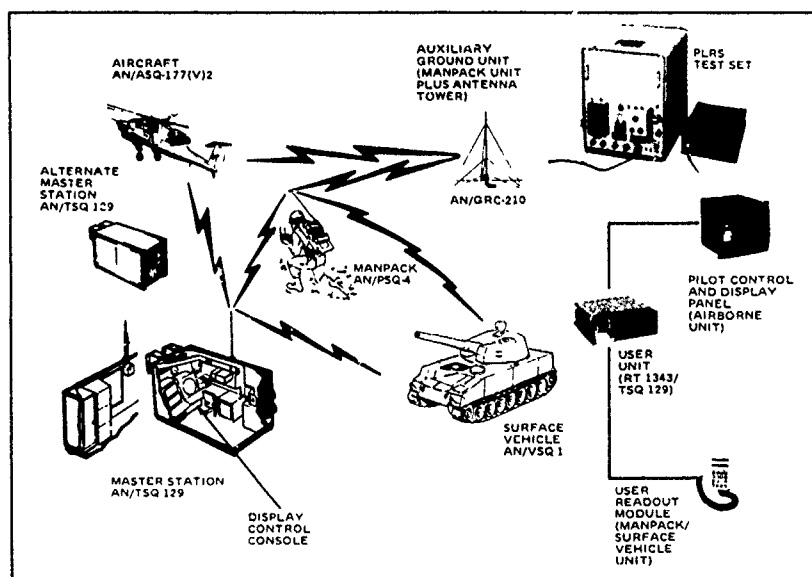


Figure 1. Major Elements of the Position Location Reporting System (PLRS).
The PLRS Master Station provides centralized reporting of position location data

The PLRS MS performs or provides the following functions: Centralized network management; automatic processing of position, navigation, and identification information for each participating user; Exchange of limited digital data communications; real-time display of users within the system's operational area of coverage; And PLRS-derived information to supported command and control centers. UUs, which are individually identifiable to the MS, perform reception, transmission (including relay), range measurement, and various signal and message processing functions necessary for position location and communication operations within the system.

A PLRS community is deployed with two identically equipped MSs, one of which is designated the alternate Master Station (AMS). The AMS monitors the MS's operation, participates in the network position location and communications functions, and assumes system control either when directed by the MS operator or automatically upon MS failure.

The PLRS user equipment is built around one key unit, the basic user unit (BUU), supplemented by unique installation kits to adapt the basic configuration to each installation type. The BUU contains the principal transmit, receive, signal processing, and message processing elements. The BUU is used in all the various UUs. Added to this unit are other installation kit equipment items to produce the different UU types.

The identification and position determination of UUs by PLRS is fully automatic; that is, when the UU operator turns on his equipment, it automatically becomes and remains a member of the PLRS network. However, to permit the UU operator to provide data and requests to the MS and to receive and display information from the MS, a separate user input/output (I/O) device is required and is usually employed with each UU. For the manpack unit (MPU) and surface vehicular unit (SVU), this I/O device is a small hand held unit called the user readout module (URO). For the airborne unit (AU), a pilot control and display panel (PCDP), is provided. The URO and PCDP are, with a few exceptions, functionally and electrically identical.

NETWORK STRUCTURE

Tactical deployments require operations beyond line-of-sight (LOS) from the Master Station. The approach taken in PLRS to satisfy this non-LOS requirement is to use relays. In some deployments two and sometimes three relay levels are needed to establish a path between a remote UU and a MS, which reports positions to the command and control center. Deployment of sufficient dedicated relays to provide acceptable coverage in a mobile environment would be expensive. Therefore an integral relay capability is built into every UU. Any UU can be utilized to maintain contact with any other UU. This reduces the need for dedicated relays and improves speed of adaption to changing deployments.

PORT Links - To maintain communications and provide organized reporting of data for position location calculation, Hughes developed a concept of organization which is called a PORT structure. This is a communications structure (see Figure 2) consisting of a set of PORT links which connect UUs (nodes) to the MS either directly or via one to three relay nodes. Directly connected units are termed "A" level nodes; units connected via one relay are called "B" level nodes, etc. Thus a "D" level unit corresponds to a node for which the path contains four PORT links and three relay nodes (one A, one B, one C). The "D" level is the highest level required by the operational PLRS. Pairs of TOA measurements are made on every PORT link and reported back to the MS. Thus, in addition to range, a UU's clock can be tracked using paired TOAs. "A" level clocks are regularly checked against the master, "B" level clocks against the "A"s, etc.

TOA Links - In addition to the bilateral PORT link, one-way TOA links are utilized to provide the additional multilateration structure needed for position location and tracking. Since the timing of UU clocks is established using paired TOA data along the PORT paths, then one-way TOAs can be converted to range estimates. In a typical PLRS deployment over half of the range measurements are based on one-way TOAs.

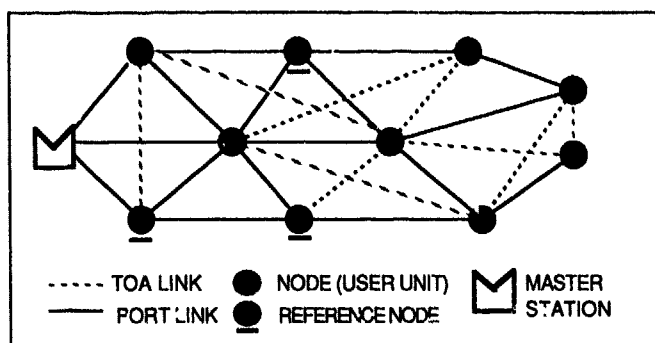


Figure 2. Simplified PLRS Network Structure. Network control and measurement reporting is transferred over the PORT path.

Reference Nodes - In order to initialize the position location function and to maintain a relationship between the internal coordinates and the external military grid reference (MGR) coordinates, the MGR positions of three or more cooperating units are input to PLRS. These are normally input as three dimensional fixed reference positions and the units become full reference nodes. The Master Station may be one of the reference units. Unit positions may be entered as fixed altitude or fixed horizontal reference units, which allows flexibility in special cases. In addition the system can operate without any fixed reference units as long as the positions of three or more units are regularly input to the position tracking function. In this latter case the positions may be input and updated by units which are moving. This is termed a dynamic baseline operation.

POSITION LOCATION AND TRACKING

The initial location of position within PLRS is based on the use of three ranges and an altitude to unambiguously locate a new unit in three dimensions (see Figure 3). The altitude, based on a barometric measurement, establishes the unit on a horizontal surface; two ranges, based on TOA measurements, are then used to establish a pair of points on that surface; and the third range is used to resolve the ambiguity. Velocity is then established and confirmed using a sliding three point method for filter initialization. All of the position locations within PLRS are established and tracked at the MS based on measurements taken by the UUs. This data is provided to the MS via user measurement reports. Each user measurement report message may contain up to three TOA measurements and a barometric measurement.

Choice of an algorithm for position location update is strongly influenced both by the user unit dynamics and by the multi-level relay requirements. The majority of user units are not LOS to the Master Station and in a large network it is unlikely that most units have LOS to any known reference units. Thus a typical unit must be located by using measurements from other units which are being tracked. Since all position measurements cannot be made simultaneously, it is necessary to extrapolate the position of one (or both) units cooperating in a range measurement. In the PLRS approach, a portion of a position correction usually is ascribed to each unit involved. The amount of correction applied is a function of the track uncertainty of each unit along the line-of-path between the two units.

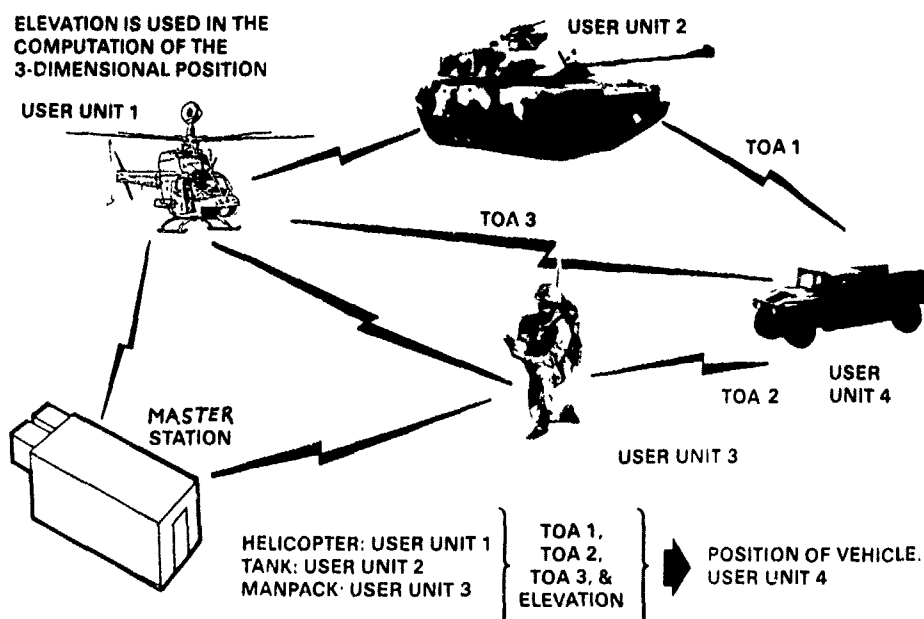


Figure 3. Position Location in PLRS. A combination of TOA and barometric measurements are used by the Master Station (MS) to locate and track user units (UUs)

Because of the availability of multiple links and their usefulness in tracking other units, a single unit may have up to ten or more TOA measurements taken at various times, utilized in a single position report period. To maintain a simple algorithm which adapts to the variable data base and minimizes computer memory requirements Hughes chose to process each TOA as it is received at the MS. This provides a sequence of partial updates and makes the tracking algorithm relatively independent of which unit is the primary beneficiary of a given TOA.

TRACKING FILTER OVERVIEW

There are four, adaptive, predictor-corrector filters used in PLRS. All of these filters are simplified versions of the discrete Kalman filter. They are implemented in the software of the MS's TOA processor, an AN/UYK-7 computer. Together the filters take the raw measurement data from each UU's measurement report message of TOA (time-of-arrival) and barometric pressure transducer values, and convert them to updated estimates of the user's three-dimensional position and velocity. Figure 4 depicts the interconnectivity of the four filters. All filters take one piece of input data at a time.

The mean sea level (MSL) filter's purpose is to furnish an offset calibration for all the non-reference UU barometric pressure transducers by using reference UU barometric pressure transducers data as input (reference UU are at known, fixed altitudes).

The output of the MSL filter is used as a constant by the altitude filter. The purpose of the altitude filter is to obtain vertical position and vertical velocity estimates based on barometric pressure transducer input when there is little vertical information in the TOA data. The altitude filter output is also required to aid position locating and checking of entries (PLACE) in initialization of the track review and correction estimation (TRACE) filter. PLACE is an algorithm, rather than a filter, which provides initial track acquisition, ambiguity resolution whenever possible, and initial position and initial velocity estimation for TRACE.

The central logic oscillator control (CLOC) filter is used to obtain estimates of each UU's clock offset and drift rate with respect to the MS's clock (or oscillator). With this knowledge, precise one-way TOA ranging is possible without using prohibitively large and expensive highly stable oscillators at each UU. CLOC provides, as required, commanded corrections to each UU's clock offset and/or frequency to keep all UUs nominally synchronized with the MS, and thereby with one another.

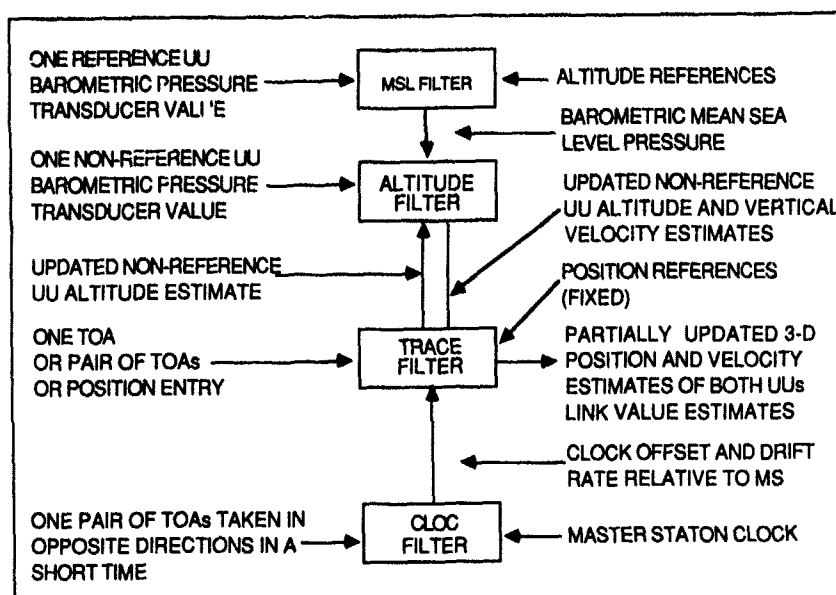


Figure 4. Tracking Filters Interconnectivity and Data Sources.
One piece of input data is taken at a time.

Finally, the purpose of the TRACE filter is to take each one-way TOA and partially update the two cooperating UU's position and velocity estimates in three-dimensional space. Without the supporting processing from the other three filters, TRACE would not be able to do its job so successfully. Link value accounting is one of the unique fallouts from the TRACE processing. If a particular TOA link assignment is not aiding the position location accuracy (due to geometry or excess TOAs in a particular direction) of either UU then that TOA link assignment is replaced by one that may be more beneficial to overall system accuracy.

These filters provide the partially updated positions necessary to permit the MS to report a fully updated position (if desired) about 1/2 second (2 frames) after MS receipt of each user's measurement report.

WAVEFORM ARCHITECTURE

PLRS is a fully synchronous time-division multiple-access system that employs a spread-spectrum waveform to perform multilateral position location and tracking. There are other radio systems that provide position location. There are others that send and receive digital messages. And there are still others that provide a large degree of rejection to jamming. The PLRS differs from nearly all of these in two fundamental aspects.

One significant departure is that PLRS employs spread spectrum, with many units in a network time-sharing a single channel. Since two units seldom carry out their respective function on the same frequency in the same instant of time, any latent problems associated with high power stations "taking over" the link are avoided. This may be contrasted with more conventional direct sequence systems that transmit continuous signals and rely only on code division multiplexing to provide compatible transmit/receive operations.

Employment of this synchronous time division spread spectrum technique is greatly facilitated by use of a digital matched filter. The matched filter is based on an integrated circuit that allows a large segment of a received signal to be examined for synchronization simultaneously. Since the receiver can examine a large number of code bits (chips) at once, rather than having to try them one at a time as in a conventional spread spectrum receiver, the amount of synchronization time required is greatly reduced and one of the prime objections that has prevented widespread use of spread spectrum systems in the past is avoided.

Another major departure from prior systems is that PLRS employs a network which is fully synchronous in three respects, as follows: all units maintain timing such that crypto resynchronization is seldom required; all units perform actions in a programmed cyclic manner such that reprogramming UU assignments for relay, ranging, and reporting is seldom required; and each UU's time base is maintained with sufficient accuracy, such that one way time of arrival measurements can be translated to ranges by the MS. Each of these aspects reduces the number of required transmissions and makes more time available for other system functions or for increased system capacity.

The basic concepts involved in PLRS (the spread spectrum waveform; synchronous implementation of time division multiplexing, and multilateration tracking) are discussed in the following paragraphs.

SYNCHRONOUS TIME DIVISION MULTIPLEXING

PLRS employs time division multiplexing to permit a large number of users to utilize the same frequency (Figure 5). Each of the PLRS units in a network takes turns transmitting its burst while other units listen. These time slots are assigned by the Master Station based on the particular requirements of each user. (For a given tracking accuracy, an aircraft-mounted unit needs more time slots than a manpack unit because of its higher dynamics.)

In the synchronous TDMA approach, every user unit has its own time base generator that keeps track of the time that it should transmit or perform other programmed operations. Once this time base is synchronized with the Master Station's time base, then messages can be sent or received and range measurements made. The Master Station's time base acts as the prime timing source for the entire network, correcting each of the user unit clocks whenever they require it. In this way, the timing oscillators included in the user units need to have only moderate stability, and can be inexpensive production type components.

The main advantage of fully synchronous operation is that more time slots are made available for network users, although there are other benefits that accrue. Since the Master Station does not have to contact a user unit each time it wants it to transmit for ranging, a great number of time slots (nearly half) are freed for other uses or users. This capability to provide full net operation with fewer transmissions also is used to help reduce interference with other systems; to reduce the probability of signal detection; and to increase the probability of successfully completing critical transactions in a jamming environment.

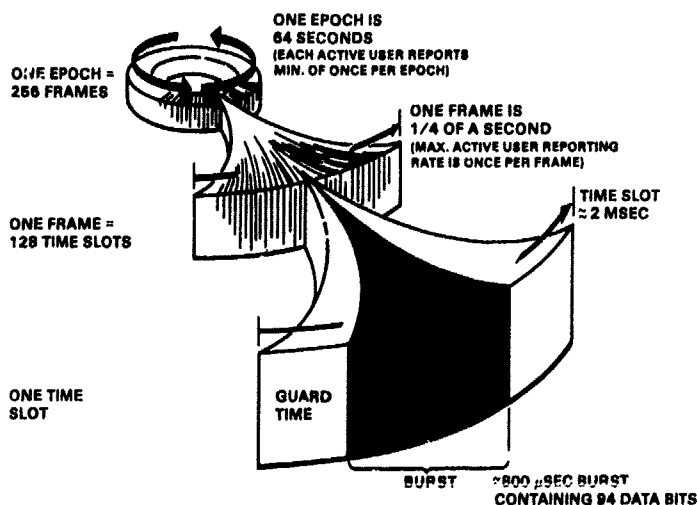


Figure 5. PLRS Time Division Multiple Access (TDMA) Organization

RANGE MEASUREMENTS

One-way ranging is made possible by the fully synchronous nature of the PLRS network. Each user unit employs a set of time markers to designate when a transmission is to start and when a reception must be completed. Thus a user unit needs only to measure the time delay from the end of the reception to a time marker at the end of the time slot. This yields a digital number (TOA) precisely related to the range between the two units. (See Figure 6)

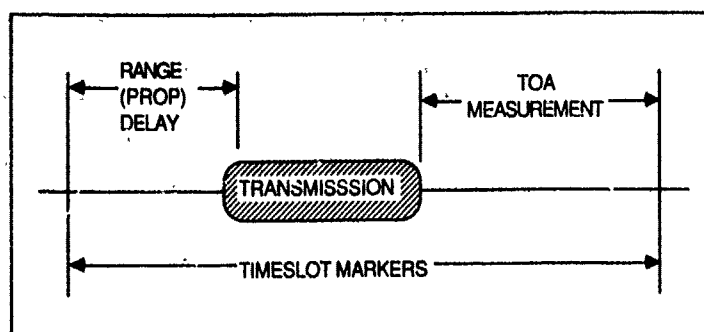


Figure 6. Synchronous Ranging. The TOA is precise measure of range delay.

Spread spectrum signals are especially amenable to range measurement because of the high chip rates employed in their modulation and because those very high chip rate code sequences have excellent correlation functions. Over the entire length of a code sequence, there is only one point at which a code will correlate with itself, and that point is only about one chip increment in length, and standard methods have been worked out for spread spectrum systems that allow them to resolve range to a small fraction of a bit.

To provide time synchronization, the system performs time difference measurements (Figure 7). When the MS transmits, a user unit measures the TOA. When the user unit transmits, it reports its TOA measurement. The Master Station uses that information, along with the TOA information from the user unit's clock offset. When required the MS tells the user unit to correct its timing. All timing information is stored in the Master Station's computer, along with position information derived from TOA measurements.

SPREAD SPECTRUM WAVEFORMS

PLRS performs its functions in the face of either deliberate or accidental interference. One of the design features that makes this possible is the use of a spread spectrum type signaling waveform. Specifically, the information transmitted is spread to a bandwidth of approximately 3 MHz by modulation with a pseudonoise code sequence. Each time a user unit or the Master Station transmits a burst, that signal burst is spread by the code. The spread spectrum signaling format provides a low density signal spectrum that reduces detection by would-be interceptors and offers minimum interference to other co-channel users. In addition, the effect of this modulation is to encode the signal and thus to help protect it from those who might try to extract information if the signal is detected.

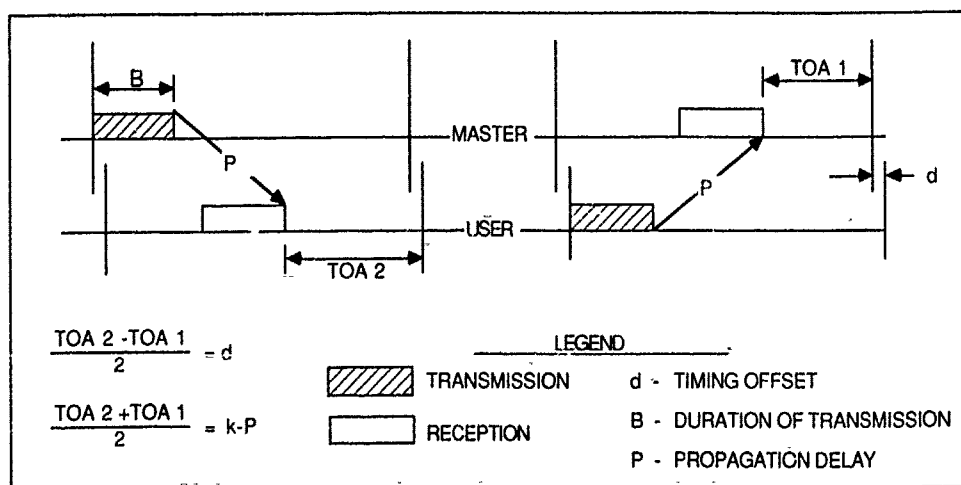


Figure 7. Time Difference Measurement Technique. Since all timing information is stored in the MS Computer, the MS can predict future variations by extrapolating previous drift rate.

At the receiver, the signals that arrive (a composite of desired and interfering signals) are all compared in a correlator with a local reference code, and the incoming signal components that match the local reference are selected. Since the reference is a wideband replica of the desired signal, the effect of the correlation process is one of turning the desired wideband input signal into a narrow band signal and causing undesired input signals to be spread out over the local reference bandwidth or wider. For example, if the undesired input is from a CW jammer, then at the correlator output it is a replica of the local reference; if it is a wideband signal such as from a noise jammer, then it is a composite of the noise and the wideband signal (i.e., it comes out much wider than it goes in).

The receiver, knowing the post-correlation bandwidth of the desired signal, employs a narrowband filter that can just pass the re-mapped correlator output. Any signal except the desired synchronous signal is spread into a wide bandwidth at the same time the desired signal is caused to "just fit" the post-correlation filter. Hence, almost all of the undesired signal will lie outside the passband of the filter while the desired signal passes neatly through. A demodulator then views a desired signal that has been greatly enhanced with respect to the interfering signal.

The burst of signal that is sent by a PLRS unit consists of two portions - a preamble portion and a message portion. The preamble portion is made up of code bits (chips). A PLRS receiver examines the chips, applying them to the digital matched filter that accepts or rejects the signal on the basis of the degree of correlation between the chips received and those expected. If the preamble is accepted, then the message, which consists of addresses, commands, queries, and replies to queries, can be decoded.

All of the burst signals appear to be identical. That is, whether the particular signal being sent is a reply to a query, a request for information, or whatever, it has the same structure and cannot be distinguished from any other signal without having the proper receiver and crypto key. Also, ranging is accomplished with any of the bursts sent, no matter what their purpose as far as the message portion of the burst is concerned.

NETWORK ADAPTATION TO THE ENVIRONMENT

Each assigned TOA link has an associated quality which is the combination of its reliability and its "value" to position location. When a link's value changes (either due to a unit's motion, propagation effects or ECM), the change is noted and if the value drops too low, the link is replaced with another.

Within the network structure, every unit must have 1 or more PORT links. To insure this, the MS keeps track of the reliability of each link and switches links when they become unreliable. For example (see Figure 8), if PORT link b breaks for any reason, communication from the MS is lost to both units C and E. Since, however, C is known from TOAs to communicate with both B and D either link h or i may be changed to a PORT link, and no searching or testing is required. Say link h is chosen; E has a TOA with D, hence link j is known to be reliable for communications and is therefore changed to a PORT link. (Note that even though link c may be good, its use as a PORT link is dropped when unit E is assigned through D. This is done for three reasons: one, it is faster to reassign unit E to another unit than to retain it through the old PORT unit C. Two, the break in link b may have been due to unit

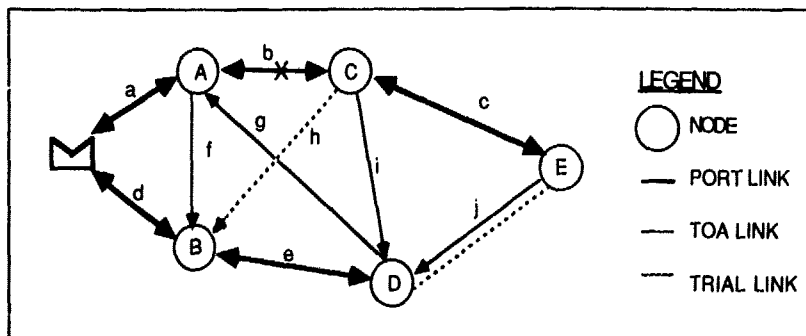


Figure 8. PLRS Network Adaptation. Measuring reliability and value on each link allows the MS to adapt the network to the changing environment.

C. (Unit C may have moved behind a hill which can also show up as a break in link c or it may have been turned off or destroyed.) And three, to diversify the network, that is, to allow the network to find its own preferred configuration aided by use of measured link reliabilities. By measuring the link reliabilities, the network can adapt itself to changes in the environment by automatically re-routing through known good links.

PLRS NETWORK MESSAGE TRAFFIC

Four types of messages may be transmitted through the PLRS network: user data output messages; user data input messages; network control messages; and measurement report messages. All messages are sent to their appropriate destination, through relays if necessary, without data content change. Error control is used on each to insure low error rate.

User data output messages originate at either a command and control center through an interface to the PLRS MS or at the MS itself. This message is sent through the network to the appropriate user unit and its input/output (I/O) device. An I/O device may also be used to originate queries and data messages to be sent back to the MS. This establishes two way communication between the command and control center or MS and any user unit.

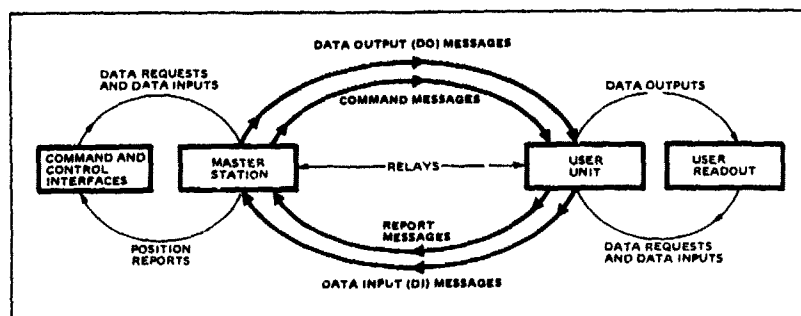


Figure 9. Link Message Flow. Four basic link message types provide internal PLRS control and reporting, as well as limited external communications.

Network control messages and measurement report messages are used by PLRS to maintain communication with, and exercise control over, each unit. The network control messages contain commands to the user unit, (*e.g., link assignments and timing correction commands). Measurement report messages contain TOA data, user queries, altitude, and status information.

The actual routing of the message to the proper destination is implicit in the network structure. The MS, with its knowledge of the network, assigns transmit and receive times to provide proper relaying of messages to their destinations. Units performing a relay function make no distinction between different message types or, in general, the direction a message is traveling. All the unit needs to know is when to receive and when to transmit for relay.

ENHANCING PLRS WITH USER-TO-USER DATA CAPABILITY*

by
J. A. Kivett and R. E. Cook
Hughes Aircraft Company
Communications Systems Division
Ground Systems Group
P.O. Box 3310
Fullerton, California
U.S.A.

ABSTRACT

The Enhanced Position Location Reporting System (EPLRS) maintains all of the basic PLRS capabilities while greatly increasing the user-to-user data capability. The EPLRS utilizes the proven PLRS control network for monitoring and controlling large communities of user terminals including the positioning, position reporting, navigation aid, cryptographic key distribution, and status reporting functions. In addition, the control network is utilized for distributing communications circuit assignments and monitoring user-to-user communications performances.

Both duplex (point-to-point) and group addressed (broadcast) types of service are available via the same user terminal. Each terminal can support many user circuits (needlines) simultaneously with a composite (receive plus transmit) information rate in excess of three kilobits per second. The primary user interface is via the PLRS/JTIDS Hybrid Interface (PJHI), which uses CCITT x.25 protocols, but an FSK interface is also available for backward compatibility with existing systems such as TACFIRE.

The EPLRS concepts have been proven through live testing using prototype terminals, including interfacing with six military host systems. In addition, extensive computer modeling and large scale user community simulation have been used to confirm extension of performance to full scale operation with up to one thousand user terminals in a division area. Limited production of over two hundred terminals is underway, with development and operational testing scheduled to begin early in 1988.

INTRODUCTION

The Enhanced Position Location Reporting System (EPLRS) extends the original PLRS concept to include user-to-user data communications. The capabilities of the USMC/USA Position Location and Reporting System (PLRS) uses advanced Time and Frequency Division Multiple Access Spread Spectrum technology. PLRS is in the process of being put into production and EPLRS is finishing Engineering Development.

The EPLRS system will support the Army's near term requirements for data distribution, position location, navigation, and identification in a battlefield environment. EPLRS, under a five phase development plan, will go into Developmental Testing/Operational Testing evaluation in 1988 and subsequently evolve into the objective Army Data Distribution System.

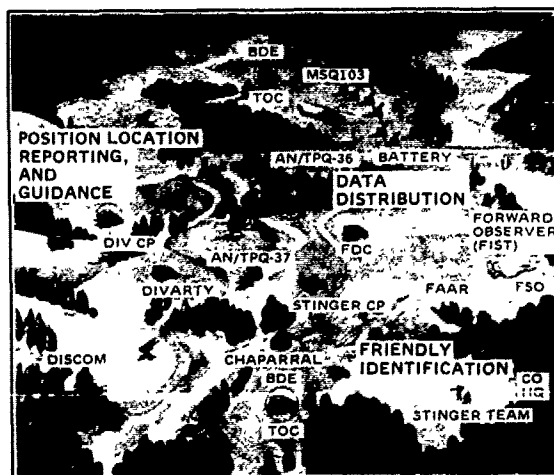


Figure 1. EPLRS Maximizes Battlefield Effectiveness

* This work was supported by the U.S. Army under contract DAAB07-82-C-J096, administered by U.S. Army Communications-Electronics Command, Fort Monmouth, NJ.

PLRS/EPLRS interoperability allows for mutual support and coordinated operations between Basic PLRS and EPLRS equipped units. Coordination of resource allocation, Areas of Responsibility, and the lateral transfer of Command and Control information are all accommodated by the architecture.

DEPLOYMENT

EPLRS System Elements - EPLRS supports all five general mission areas in the Army (i.e., Air Defense Artillery, Fire Support, Intelligence/Electronic Warfare, Combat Service Support, and Maneuver Control). EPLRS contains two primary equipment elements: Enhanced PLRS Terminals (PTs) and Network Control Stations (NCSs). The PTs are assigned to Division and Corps units that participate in data communications, identification, and position location/navigation. Between 500 and 820 PTs are assigned to each Division. In addition approximately 970 PTs and 6 NCSs are assigned to the Heavy Corps. Each separate Brigade and Armored Cavalry Regiment is assigned one NCS and from 170 to 290 PTs.

Area Coverage - A key consideration in planning for Ultra High Frequency radio communication is the line of sight propagation characteristics inherent in this band. The PT operates in the 420-450 MHz range. Modeling has shown that percent connectivity on the order of 10% is adequate. A deployment of PTs providing good area coverage is essential to take full advantage of the EPLRS system's integral relay feature. The adaptive automatic relay capability of the network allows the support of long range circuits. The proper employment of PTs deployed as dedicated relays is important to planning for and utilization of the system. Otherwise, too sparse a distribution of PTs in hilly or heavily foliated terrain may result in increased percentage of unsatisfied circuits.

Gateways - Data communications requirements which extend beyond the boundary of a Division sized area is supported by gateways. Two PTs connected back-to-back by a cable between their data ports form a gateway. One PT is in one NCS community and one PT is in the other. Information is passed between the two PTs at baseband so that each of the gateway PTs can maintain timing and cryptographic synchronization within its own community. Data communications between a source PT in one community and a destination PT in another is accomplished by setting up a circuit between the source and the gateway PT. A similar procedure takes place in the other community. Data transverses circuits within the two different communities and across the gateway, providing the functional intercommunity circuit.

SYSTEM ARCHITECTURE

The NCS establishes control circuits between PTs and the NCS via the control network. This control network supports the position location, navigation aid, friendly unit identification, and over the air rekey functions. Basic PLRS and EPLRS terminals operate compatibly using the control network, but user-to-user communications are supported solely by EPLRS Terminals (PTs). All PTs are capable of being assigned as relays in both the control and communications networks. PTs can simultaneously support control, duplex communications and group communications assignments as indicated in Figure 2.

Host-to-Host Data Communications - Host-to-Host data communications are provided via the EPLRS communications network. The communications network consists of both duplex and group addressed circuits. Duplex circuits are used to support requirements between two users requiring end to end acknowledgment control. A maximum of four PTs can be assigned to relay on a duplex circuit. Group addressed circuits are used to support requirements between multiple users not requiring acknowledgment control. Only one PT may source information on a group addressed circuit path. A maximum of three PTs can be assigned as relays on a group addressed circuit path.

Data Communication Message Flow - Communications messages are input to the system via the PLRS/JTIDS Hybrid Interface (PJHI). This is an X.25 type interface where data flows in packets containing up to 128 bytes (1024 bits) of user information. Destination packet addressing information is integral to the header of each data packet in the form of an 8 bit Logical Channel Number. Each host message is mapped into PJHI packets with a corresponding Logical

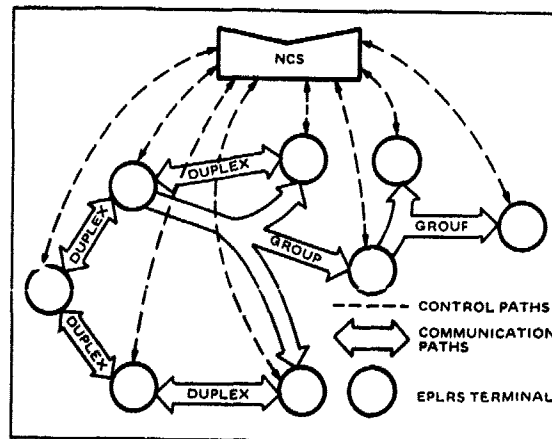


Figure 2. Control and Communications Networks in a Single NCS Community.

Channel Number by the Host so that the subsequent transmission of the message can take place over the proper circuit. The PTs convert PJHI packets into a form for EPLRS transmission called Transmission Units. Each Transmission Unit contains 80 information bits. Transmission Units are transported through the communication network via NCS assigned virtual circuits. The PTs are responsible for acknowledging and sequencing Transmission Units and circuit monitoring. The NCS assigns PTs to participate on communication circuits as sources, destinations, or relays. The Transmission Units are sent between source and destination PTs through relays, where the adaptive circuit routing is controlled by an NCS. Transmission Units are converted back to PJHI packets at the destination PT.

All control network traffic is nodal to the extent that the message routing takes place through the NCS. In contrast to the PT-to-PT traffic (via relays) occurs on the communication network, with no involvement by the NCS except to set up and maintain the circuits.

From a message flow standpoint, the control network provides an "order-wire" capability that can be used to establish and maintain the communication network. In addition to the order-wire, the control network also supports NCS services. These services provide the PT user with a wide range of position location, navigation, and identification information. The communication network continues to function in the absence of NCS control for as long as the PT deployment locations remain relatively static and a steady state environment exists (i.e., no severe jamming attack occurs).

EPLRS Circuits - An EPLRS circuit is defined by the following parameters: (1) the source and destination Military ID and associated Logical Channel Numbers, (2) circuit priority, (3) circuit rate, and (4) duplex or group addressed circuit.

The Logical Channel Numbers are used in place of addresses. Since PTs are capable of operating with multiple Logical Channels simultaneously, the assignment of Logical Channel Numbers is important so the host messages may be routed over the correct virtual circuits.

Four circuit priorities are accommodated. Circuit priority is a key software consideration in the circuit activation process. In case of heavy system loading, requests to activate low priority circuits receive delayed responses, or in extreme cases the lowest priority circuits are not implemented.

The rate selected for a circuit dictates how large a block of timeslot resource will be assigned to satisfy the requirement. This rate is determined based on both the throughput and response time requirements. The Acknowledgment control determines whether a duplex or group addressed circuit is used.

Relay - The relay PTs programmed by the NCS are responsible for relaying Transmission Units (TUs) between source and destination PTs. The relay PTs can not read or alter the Transmission Unit data content, but they do perform error correction decoding and encoding, and validation of relayed Transmission units.

EPLRS communication circuit (duplex and group addressed) timeslot assignments are structured to allow for multiple receive opportunities. The source, destination, and relay PTs have specific times to receive defined by recurrence period, start frame, and timeslot number. Duplex Circuits - Source and destination PTs transmit on duplex circuits in a periodic manner. They transmit once per Recurrence Period. On duplex circuits, the timeslots are arranged such that for each source transmission there is a destination transmission allowing for balanced communications with acknowledgment between the two PTs.

Each duplex circuit structure has at least two paths. Path 0 is used for the first transmission in the epoch and every other transmission after that, and path 1 is used for the second transmission in the epoch and every other transmission after that.

Group Addressed Circuits - As in the duplex circuit structure, the group addressed circuit structure is periodic with respect to the Recurrence Period, and the timeslots are defined by start frame and timeslot number. These parameters allow a source to know exactly when to transmit. Because the reception of messages on group addressed circuits are not acknowledged, the relays need only transmit in one direction. The timeslots are arranged so that the source can continuously transmit (up to 1200 bps). The relays and destinations on a group addressed circuit are not programmed for a specific path or level. Therefore, it is not known what path(s) will be used to relay the messages to the destinations. This approach is commonly known as path diversity broadcast.

Second Source Receive - Both duplex and group addressed circuits use the second source receive technique. This technique assigns a set of timeslots in which relays and the destination PT "listens" for the source PT's transmission and all other upstream relay transmissions. This allows PTs to receive transmissions early and increases the probability of receiving transmissions. For example, the fourth-level PT could conceivably receive a message during the transmission from the source or any subsequent relay of the message.

ENHANCED PLRS USER UNIT (EPUU)

The PT consists of an Enhanced PLRS User Unit (EPUU), a User Readout device, and an appropriate installation kit for ground, surface vehicle, or airborne use. The EPUU can be configured to support Hosts via either the PJHI or a Frequency Shift Keying (FSK) Interface. The EPUU is designed to meet the environmental requirements of ground and airborne installations. The User Readout device serves as a control panel for the PT, and for handheld use with ground based installations. The data access capabilities of the User Readout is limited to short (10 characters at a time) message exchanges via the NCS.

The PT performs user-to-user communications and generates ranging information for position location/navigation. The NCS can assign any PT as a relay of opportunity. This extends the communications range and keeps communication links operable during jamming or changes in user location.

Data Flow - The key data flow steps in the EPUU are shown in Figure 3. After stripping the user specific protocol and replacing it with the EPLRS network protocol, each Transmission Unit is encrypted with a user variable, then a (104/94) Cyclic Redundancy Check error detection code is applied. After this, the Transmission Unit is transmission encrypted, and error correction encoded. Encoded blocks are interleaved to protect against burst errors. Finally, each channel bit is modulated by a pseudo-noise code for spread spectrum protection and the Transmission Unit is frequency hopped to provide increased antijam protection. At the receiving terminal the process is reversed.

The EPUU is a multifunction radio that generates and processes EPLRS messages. Centralized control of the EPUU by a microprocessor within the Message Processor allows partitioning of processing into logical EPUU functions as shown in Figure 4.

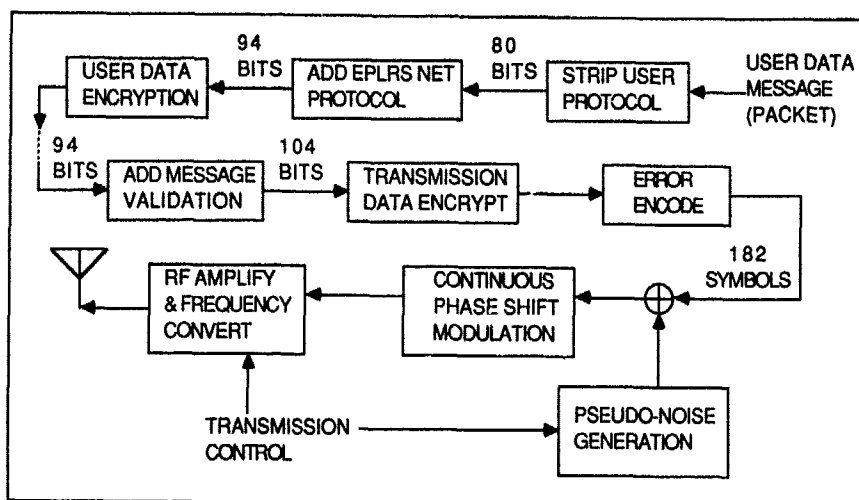


Figure 3. PT Data Transmission Process

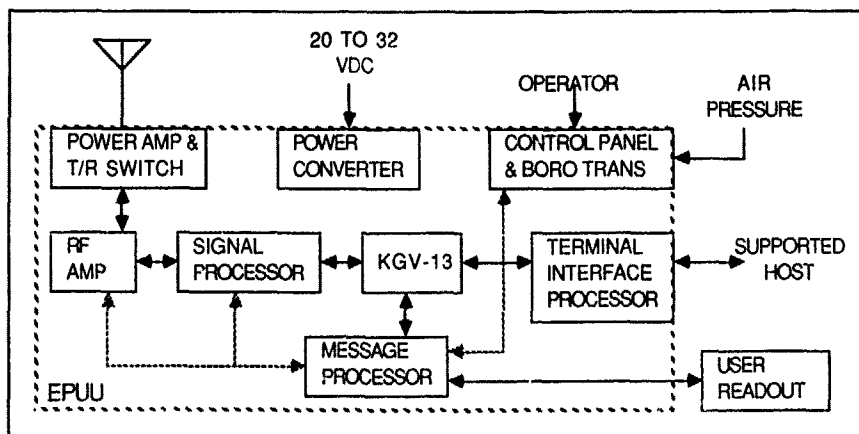


Figure 4. Major PT Functions

RF/IF Function - The RF/IF function performs frequency conversion, amplification, and filtering of the transmitted and received signals. During receive, this function performs a 2 BIT A/D conversion of the incoming signals. During transmit, the digital output of the Signal Processor is used to generate the Continuous Phase Shift Modulation.

Signal Processor Function - The Signal Processor function performs preamble detection/generation, interleaving/deinterleaving and error correction/error detection encoding/decoding, pseudo-noise code generation, data correlation, and Time-of-Arrival measurement.

Secure Data Unit Function - The Secure Data Unit function is performed by the KGV-13 which encrypts/decrypts transmitted and received data, performs message validation, and provides outputs for transmission security.

Message Processor Function - The message processor contains a CMOS microprocessor that is the central controller of the EPUU. This function controls all processes done within the EPUU. The message processor, additionally, generates and decodes link messages.

Terminal Interface Function - The Terminal Interface Processor provides the interface to the PJHI link or the Frequency Shift Keying interface. The PJHI has as its objective the formulation of a standard interface to which all tactical data systems can design. This interface defines packet generation for data distribution in a rigorous and logical manner through the first three levels of the International Standards Organization (ISO) model. The Frequency Shift Keying interface is capable of supporting a 1200 baud channel between the Frequency Shift Keying tactical data system and the PTs.

NET CONTROL STATION

The NCS is a tactically packaged computer facility providing centralized technical control and monitoring of an EPLRS community. The NCS performs dynamic network management of all PTs under its control as indicated in Figure 5. The NCS is the central control point of the PLRS network, allowing a technical control operator to maintain a real time overview of the network within his area of responsibility.

The NCS software operates in real time. The NCS's operational software provides centralized network management and automatic distribution of position, navigation, and identification information.

The terminals accept and implement NCS issued commands and report status and communicant data (i.e., who can hear whom). These reports are essential for accomplishing the automatic adaptive routing capability of the EPLRS system.

NCS Operator - The principal human interface to the system is at the NCS. The NCS operator, as the communications technical controller, is responsible for network initialization and monitoring ongoing community performance. As PTs enter the network, the NCS operator monitors the performance of the network using a graphic display. Normal operation requires only minimal operator intervention. His work is intended to be primarily exception processing, taking actions to help the automated network management, and to apply data access and resource management guidelines provided by System Control (SYSCON).

Network Management - The Network Management software functions automatically provide for dynamic network monitoring and resource assignment. In the battlefield environment, the operating radio links (connectivity) available in the network can be very limited and rapidly changing. There are two design features of the EPLRS system that enable it to provide continuous service in such an environment. First is the high antijam capabilities of the spread spectrum EPLRS radios, without which there would be very little connectivity in a jamming environment. Second (and just as important in the dynamic environment) is the automatic network management function which controls the EPLRS network configurations in real time. The network management function within the NCS performs dynamic, real time communications routing and timeslot assignment to the PTs and monitors and controls the subsequent network performance.

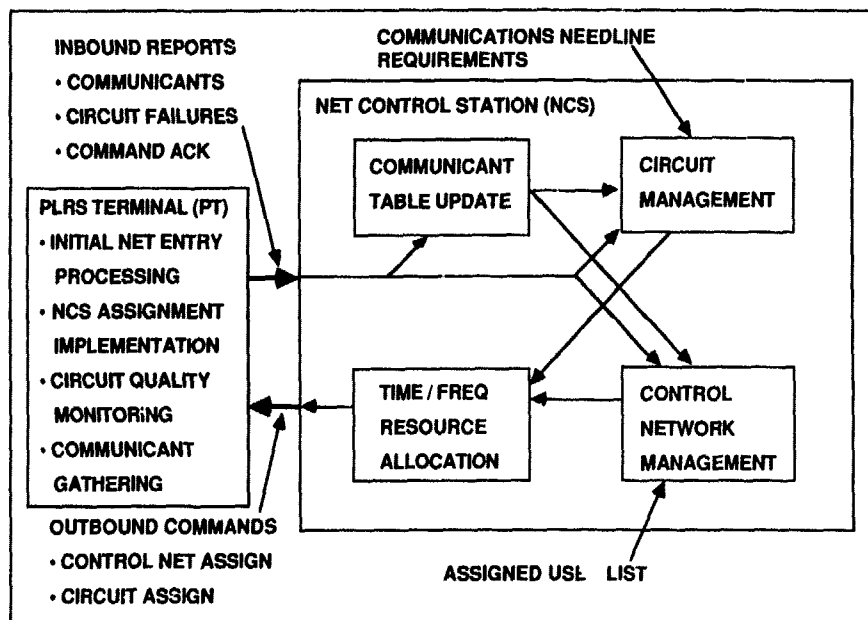


Figure 5. Major NCS Circuit Management Function

The communication network management functions are distributed to both PTs and NCS. The PTs communication network management role is to implement the commands issued by the NCS, detect circuit failures, and control the flow of data packets into and out of the network. The NCS uses the control network to deliver the commands used to build and maintain the communication network. Communication network management consists of three basic functions. They can be generically categorized as: Message Traffic Control, Circuit Management, and Circuit Assignment Selection.

Message Traffic Control Function - The main responsibilities of Message Traffic Control for communication network management are to receive/distribute status and measurement reports and to route, schedule, and control the acknowledged delivery of the commands sent to PTs.

Circuit Management Function - The main responsibilities of Circuit Management are prioritized queuing of circuit requests and maintaining circuit status.

Circuit Assignment Selection Function - The Circuit Assignment Selection finds paths and resources that satisfy the circuit requirements. The paths are found using a minimum cost path finding algorithm. Circuit assignment or reassignment requests are processed on a priority basis. Requests with equal circuit priorities are selected on a first-in, first-out basis. Successful attempts at finding circuit assignment results in generating Communication Circuit Assignment commands for the PTs selected to participate on the circuit.

Crypto Management Function - An additional NCS function is the generation and over the air distribution of cryptovariables to all PTs. The EPLRS security architecture provides the system with both SECRET and CONFIDENTIAL cryptographic data encryption.

HOST INTERFACES

EPLRS implements a Virtual Circuit approach to satisfy host-to-host data communication requirements. A wide range of information rates and response times are provided in support of both duplex and group addressed requirements.

The EPLRS uses the PJHI (X.25 type) protocol, in which hosts input data into the communication network on Logical Channels in the form of packets via the standard PJHI. The data delivered through the network is transparent to the host. X.25 is an International Telephone and Telegraph Consultative Committee (CCITT) standard three-layer protocol and is virtual circuit oriented. PTs can participate on multiple simultaneous Virtual Circuits using different Logical Channel Numbers.

User to user data flow is initiated by the source which converts the message into one or more PJHI packets. The packets are then delivered to the PT's Interface Processor where they are converted into a form suitable for radio transmission. The data is then transmitted in the form of Transmission Units to the destination PT. Relays may be employed. The Transmission Units are then converted back into PJHI packets and delivered to the destination. The destination then sequences the PJHI packets into a message identical to the one that the source sent to its PT.

PJHI utilizes packets to organize the efficient exchange of information across the interface. Messages which are going to be transferred on the PJHI are first broken into segments, called data packets. The maximum size of a PJHI data packet is 128 bytes. These data packets are transferred across the PJHI independently allowing messages to several destinations to be time multiplexed. This independent transfer of packets increases efficiency of information flow.

The PJHI is a functionally layered set of communications protocols implementing the bottom three layers of the International Standards Organization (ISO) Open System Interconnection (OSI) reference model. The three layers implemented by X.25 and PJHI are the physical layer or level 1, the link layer or level 2, and the packet layer or level 3.

Physical Layer - The PJHI physical layer provides the media to transfer bits from one physical device to another. The physical media in the PJHI is either NRZ digital baseband or conditioned di-phase.

The NRZ digital baseband interface may be utilized for devices that are located within 5 meters of the PT. This interface consists of six signal pairs, two pairs each of data, clock, and control. This interface is similar in design and function to the Electronic Industries Association standard RS-449/422. The conditioned di-phase interface may be utilized to remote the PJHI devices up to 1000 meters. This interface requires only two signal pairs and a shield.

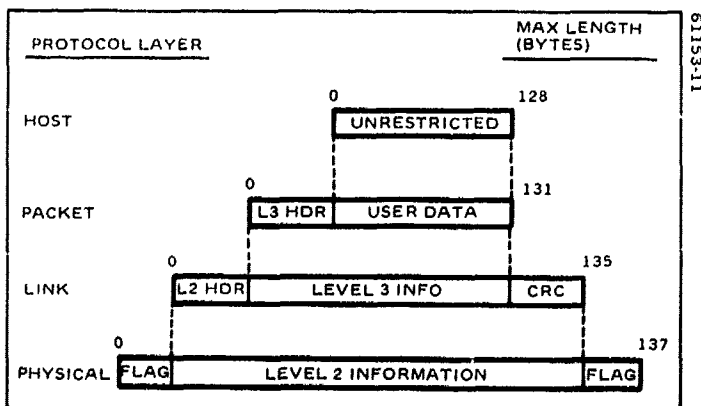


Figure 6. Information Units in PJHI

Link Layer - The PJHI link layer provides the functional and procedural means to maintain a connection to another PJHI device while detecting and possibly correcting errors which may occur in the physical layer. The PJHI performs the X 25 link level protocol by utilizing the Link Access Procedure B and the High Level Data Link Control procedure. The unit of information transferred at the link layer is called a frame. PJHI supports three general classes of frames; Information, Supervisory, and Unnumbered. The function of the information frame is to transfer sequentially numbered frames containing information for the packet layer. Information frames are variable in size depending upon the amount of information they contain. The maximum length of an Information frame is 131 bytes. The function of the Supervisory frames is to control supervisory link functions such as acknowledging, requesting retransmission, or temporarily suspending transmission of Information frames. Unnumbered frames perform the function of supplementary control of the link such as setting up, disconnecting, or stating status of the link.

Packet Layer - The PJHI packet layer defines the protocol that routes packets from one host system to another. To efficiently control packets to several hosts, the PJHI utilizes the concept of a logical channel. A logical channel is a virtual connection from one host system to another. When a logical channel is routed through a network, such as EPLRS, the Logical Channel Number has only local significance.

In the PJHI packet layer, the transmission of data is controlled separately for each logical channel in each direction and is based upon acknowledgments from the receiving PJHI. To perform this flow control protocol, the PJHI sequences packets and defines a window protocol. Data packets are sequenced by attaching to each data packet a sequentially numbered tag. Data packets can be transferred across the PJHI when the data packet is within the window. If a data packet is outside the window, then the packet will be rejected. The receiver can convey authorization for additional data packets by incrementing a packet receive sequence number. In this way the receiving PJHI packet layer can control, on a logical channel basis, the flow of incoming data packets.

The PJHI packet layer also performs error control for recoverable errors and error diagnosis for unrecoverable errors at the packet layer. The packet layer performs error recovery by resetting the affected logical channel and starting again.

WAVEFORM

The EPLRS network resource is organized into a Time and Frequency Division Multiple Access structure. Each member of a community is assigned one or more timeslots in which it can transmit a burst while other PTs receive. To accomplish this, each PT must possess a clock that is synchronized to the clock of the NCS and other PTs.

Basic Architecture - The development of EPLRS is based on synchronous Time Division Multiple Access and spread spectrum technology. Integral cryptographic security, error detection and correction coding, and frequency hopping are employed to provide robust communications in a hostile Radio Electronic Combat environment. EPLRS provides significant performance advantages including rapid response times and effective data throughput. Communications

Security, resistance to Electronic Countermeasures/Electronic Support Measures, integral automatic relay, transmissions for ranging measurements, and freedom from voice/data contention.

Resource Definition - The network utilizes the resources of time, frequency, and code to multiplex the many operations necessary for system operation. The structured use of time allows a convenient and efficient method for gathering Time-of-Arrival data, and for managing the multiple relay levels within the network. The use of the frequency resource provides additional anti-jam protection and allows for non-interfering, coordinated operation to increase system data capacity.

The Time Resource - Each PT is commanded by the NCS to perform specific transmissions and receptions at specific times. The time division structure simultaneously accommodates both minimum and maximum network access requirements. For position location and control, these vary from manpack PTs requiring update rates approximately once every minute to high performance fixed-wing aircraft with desired updates 30 times per minute. For communications, these typically range from field artillery circuits needing two source transmissions per second to maneuver control circuits needing one source transmission per 32 seconds.

Epoch - The epoch is the longest time division which is significant to the network. The longest period for which a PT can be programmed to repeat its assignments is once every 256 frames or once per 64 second epoch. The

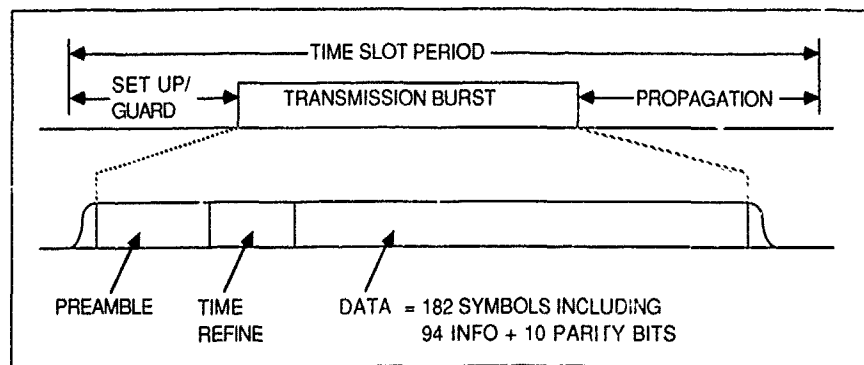


Figure 7 EPLRS Timeslot Signal Structure

programmed cyclic transmission and reception opportunities do not change from epoch to epoch unless the NCS alters the PT's assignments.

Frame - The frame contains 128 timeslots and has a period of 1/4 second. The significance of the frame is that it represents the shortest period at which a PT can be programmed to participate in the Control Network. On communication circuits, PTs can be assigned up to four transmissions within a frame on a single circuit.

Timeslot - The fundamental time division is the timeslot. See Figure 7. The timeslot length is 1.95 milliseconds. The burst transmission accounts for 800 microseconds and 600 microseconds is allocated to RF propagation delay. The remaining time is required for processing overhead such as message encoding, validation, and guard time.

The Frequency Resource - Each transmission occurs on a particular frequency channel. When operating in the hop mode, each channel is hopped across the 420 to 450 MHz band. Frequency hopping provides the network with additional Antijam performance. There are eight channels in the EPLRS, thereby increasing the capacity of the network by allowing simultaneous transmissions to occur with a low probability of mutual interference.

The Code Resource - In addition to the time and frequency resources, EPLRS uses a pseudo-noise code resource. This code resource provides different codes for each circuit in the communication network. These codes are used to determine the spread spectrum pattern used. In this way, a third dimension is added to the network capacity.

SYSTEM CAPACITIES

Network Capacity - A typical brigade sized EPLRS community consisting of 250 PTs has a nominal capacity of almost 100 Kbps in support of duplex circuits, plus a nominal capacity of almost 100 Kbps in support of group addressed circuits.

NCS Capacities - The unit libraries at the NCS define the participating PTs and their communication requirements. A maximum of 999 PTs can reside in the unit library. A nominal NCS community consists of 250 PTs and 375 circuits. An NCS is capable of supporting a maximum community size of 460 PTs. The Division wide library is capable of storing up to 1500 circuit requirements (needlines).

Circuit Capacities - The PLRS circuit capacities more than satisfy the highest rate EPLRS users as defined in the circuit requirements found in the Battlefield Communication Review Data Distribution database. The limiting factor on circuit capacity is the ability to realistically manage the circuit resources with adaptive relaying in a large network. Duplex circuits are capable of supporting acknowledged data rates of up to 600 bps. Group addressed circuits are capable of supporting non-acknowledged data rates up to 1200 bps.

Terminal Capacities - The capacities of the PTs in the EPLRS system are limited by the number of circuit assignments they can store and the timeslot allocation. A PT can support a maximum of 60 circuit assignments (30 source/destination, and 30 relay). The maximum total duplex and group addressed circuit rates assignable to a PT are 1350 bps and 1200 bps respectively.

SYSTEM PERFORMANCE

The EPLRS system is required to build to support eighty five percent of required circuits within the desired time durations. These time durations account for up to ten percent mobility of the total PTs. For this discussion, mobility refers to the fraction of PTs in motion at each instant, not the fraction capable of motion. The normal EPLRS community size refers to a Brigade Area of Operations containing up to 250 simultaneously active PTs under control of a single NCS. The initial build of both the control and communication networks for the normal community size can be achieved within 17 minutes under Electronic Countermeasures threat conditions.

Under Electronic Countermeasures conditions, the EPLRS system continues to provide communications and position location services for PTs retaining control and communications connectivity. Ninety percent of the PTs which experience an interruption in service due to Electronic Countermeasures conditions are detected and reported to the NCS within 120 seconds of the onset of the interruption. The EPLRS system then automatically repairs the communications and control paths of these PTs.

For all circuits above 150 bits per second, the Network Management function allocates the resources of the timeslot architecture and available relay PTs to achieve a one second response time for short messages. The major component of the response time is access delay, which is based on the circuit's recurrence period. The other components of median response time include source encryption, relay delay, destination decryption, signal processing, and miscellaneous data transfer delays (e.g., the PJHI data transfer).

OBSERVABILITY OF RELATIVE NAVIGATION USING RANGE-ONLY MEASUREMENTS

Alan M. Schneider
 Department of Applied Mechanics and Engineering Sciences,
 B-010, University of California at San Diego,
 La Jolla, CA 92093.

and

Naval Ocean Systems Center, San Diego, CA 92152

SUMMARY

A simulation tool is described which is capable of determining the observability of various fleet configurations and maneuvers in a relative navigation environment. The motion of the relative grid established by the navigation controller is explicitly modeled as a function of the errors in his dead-reckoning sensors. The simulation uses centralized, optimal processing of an extended Kalman filter. Results show observability on a good geometry, with some degradation in performance when dead-reckoning sensor errors change rapidly.

INTRODUCTION

Message-transmission and message-reception systems, based on propagation of electromagnetic energy, provide the opportunity for measuring the range between sender and receiver. This is the basis of the relative navigation concept considered in this paper [1,2].

Suppose that each vehicle in a network of users carries a terminal which performs the functions of transmitting, receiving, and timing of messages and also does the Kalman filtering which supports relative navigation. One vehicle of the net is designated as the navigation controller (NC). The NC defines a relative grid. A principal objective of relative navigation is to allow all users other than the NC to locate themselves in this relative grid. When the whole fleet operates without geodetic data, we have pure relative navigation. By making range measurements to the NC and/or other users, and by combining these measurements in his Kalman filter with data from his dead-reckoning (DR) system, each user other than the NC must determine his position coordinates in the relative grid and the orientation of the

northerly axis of the relative grid relative to the direction of north as indicated by his directional reference (i.e., gyrocompass).

A prime objective of this study was to develop a simulation tool which could be used to study the observability of any specified vehicle configuration and maneuvers. A secondary objective was to use this tool to assess the sensitivity of the relative navigation concept to changes in the parameters of the filter model.

Results of Prior Work by Others on Observability and Sensitivity

It has been recognized [3] that relative navigation is unobservable when all users in the network travel in formation, i.e., on parallel courses at equal speeds. It has also been recognized that some geometries are better than others, that is, observability will be marginal in poor geometries. The part of the relative navigation problem that may be difficult to observe in such situations is the direction of the relative navigation grid. A simple example demonstrates this.

Suppose there are three vehicles, A, B, and C, in the network of users, in any fixed relative pattern other than all in one straight line. A set of three perfect measurements of the ranges AB, BC, and AC determines the size and shape of the triangle ABC uniquely. However, the orientation of this triangle with respect to a directional reference established by any one of the three vehicles cannot be determined from this set of data.

It has been recognized that a (single) NC must move in order for the directions he defines for the relative grid axes to become observable to the other users. A stationary point (say, a land station) cannot serve, therefore, as the NC.

© 1985 IEEE. Reprinted with permission, from IEEE TRANSACTIONS ON AEROSPACE AND ELECTRONICS SYSTEMS, Vol. AES-21, No. 4, pp. 569-581, July 1985.

As used here, the term "sensitivity" refers to the degree of variability in the accuracy of the navigation variable estimates due to variation in

the values of the parameters of the noise model underlying the Kalman filter design.

A totally different approach to formulating the relative navigation problem is presented in [4].

The Study Plan

The simulation study was undertaken to answer the following questions: 1) Precisely how does the NC "define" the relative grid? In particular, what are the physical manifestations of this grid that can be detected by other members of the net? 2) How does a given user locate his own x-y position in the relative grid? 3) How does a given user decide where the nominally north-pointing axis of the relative grid is oriented, physically, relative to the north direction as indicated by his own directional reference? How do DR errors of the NC influence the motion of the relative grid? 4) Is it possible, that is, is there enough information from the available data, for a collection of users to unite themselves into a relative navigation grid established by the NC? The "available data" are the outputs of each user's DR system, plus accurate range measurements between users. 5) Can each member accurately predict the range and the physical direction (relative to his own gyrocompass) of the vector from himself to every other member? To be able to do this would seem to be the litmus test of the relative navigation concept [5]. 6) If indeed it is possible for the users to locate themselves in the relative grid, can this be done to a higher accuracy, or more quickly, in some geometries than others, and if so, what are the general characteristics of a good geometry? 7) Finally, how sensitive is the operation to values assumed for the parameters of the Kalman filter noise model? Since the noise models are generally simplifications of reality, one would hope that filter accuracy would not be overly dependent on these assumptions.

Design Choices in the Layout of the Simulation

One key difference between the simulation and practical application of a relative navigation system concerns centralized versus decentralized processing. In the simulation, centralized processing is used. This means that all measurements taken by any vehicle are assumed to be relayed to a central source for processing in one large Kalman filter for the whole network. While admittedly impractical from an operational viewpoint, the advantages of this choice for the simulation are a) the Kalman filter can be designed as an optimal filter for the given dynamics, measurements, and noise errors models; b) there is no problem of stability of the filter; c) no protocols need be established or simulated as to who can range on whom; any measurement between any pair of users is accepted for processing. By contrast, a practical realization would call for decentralized processing in which each user carries out a portion of the total filtering task, using only measurements he

himself has taken. Since unrestricted ranging between users in such an arrangement can lead to instability [5], selection rules have been devised governing who may range on whom. These rules can, and have, changed from time to time, and may still be changed in the future as new situations are discovered. Such rules would have unnecessarily complicated the simulation for the purposes defined. Decentralized operation leads to sub-optimal results. The centralized filter presented here provides an optimal standard for use as a basis of comparison.

The simulation studies pure relative navigation with no geodetic inputs. It handles up to four vehicles, each of which has a DR system consisting of heading and speed sensors. The NC defines the relative navigation grid. His directional reference defines the direction of grid north continuously over time. His position in the relative navigation grid as derived from his own DR system is correct, by definition. However, the DR errors of the NC cause the grid to translate and rotate; these motions are explicitly modeled. The errors of all DR sensors are modeled by first-order Markov processes, except that the NC's gyrocompass error is modeled by a second-order Markov process. The filter operates open loop; corrections to navigation states are estimated, but these estimates are not fed back to the dead-reckoners. It is a covariance simulation.

The following assumptions were made to simplify the analysis: flat Earth, two-dimensional (all vehicles at zero altitude), perfect and synchronized clocks, no ocean current or wind, no sideslip, and the statistical model of the sensor error behavior agrees with reality (no mismatch).

Filter Design

The simulation uses the extended Kalman filter, discussed in [6, pp. 182-188], modified for this application. The n-vector x , to be estimated, will be a vector of small quantities called "corrections" or "microstates." The outputs of the DR systems will be large quantities (like nautical miles east and north of the grid origin) called "macrostates" and denoted by the N-vector X . (This paper uses capital letters to denote macrostates and lowercase letters to denote microstates, with the exception that U and V , defined in Appendix A, are microstates. The prime denotes "indicated value" of a navigation variable.) The true values of the macrostates are denoted by the N-vector X . The filter is designed to be a "complementary filter" [7], in which the DR system outputs are regarded as defining the (nonlinear) navigation "process" rather than being "measurements" in the Kalman filter jargon; the filter "measurements" are the range measurements. If the true value of x were available, say, from a perfect filter, it could be combined with the indicated macrostates X' to get the true macrostates X , using the nonlinear $(N \times 1)$ vector function p :

$$X = p(X', x). \quad (1)$$

With only estimates \hat{x} of the correction states available, we obtain (estimated, improved) values \hat{X} of the macrostates as follows:

$$\hat{X} = p(\hat{X}', x). \quad (2)$$

Because there are errors \tilde{x} in the estimates of x , we will make errors \tilde{X} in the estimate \hat{X} of X .

Denoting the (discrete) measurement of range between two vehicles at time t_k by the scalar z_k , and the true range by the scalar function $h(X_k', x_k)$, we allow for the possibility of measurement error ν_k by writing

$$z_k = h(X_k', x_k) + \nu_k, \quad \nu_k \sim N(0, R_k) \quad (3)$$

where the notation $N(0, R_k)$ signifies that ν_k is assumed to be a normally distributed discrete random variable with zero mean and variance R_k .

The linear dependence of the measured range on the correction states is found by expanding h in a Taylor series around $x_k = 0$ and dropping the higher order terms.

$$z_k = h(X_k') + \left. \frac{\partial h}{\partial x} \right|_{x_k=0} x_k + \dots \quad (4)$$

The $(1 \times n)$ matrix of partial derivatives of h is denoted by H . That is, at any time t_k ,

$$H_k \equiv \left. \frac{\partial h}{\partial x} \right|_{x_k=0} \quad (5)$$

This H -matrix is the observation matrix of the extended Kalman filter. Its form is the current application is derived in Appendix B. The complete equations for the filters are given in Table I, adapted from [6, p. 110] to use the notation just developed. Filter notation generally follows that of [6]. A convention used throughout this paper is that a true state is equal to the indicated state

7

plus the correction. Thus, if X_i is the true value of a macrostate, X_i^i is its indicated value, and x_i is the true value of the correction, then by definition,

$$X_i = X_i^i + x_i. \quad (6)$$

An error is the negative of a correction. Thus, if x_i is the true value of a microstate, \hat{x}_i its estimated value, and \tilde{x}_i the error in the estimate, then by definition,

$$x_i = \hat{x}_i - \tilde{x}_i. \quad (7)$$

The Role of the Navigation Controller in Defining the Relative Navigation Grid

Many grids used in navigation are fixed to the Earth. The relative navigation grid, although intended to be fixed, and ideally so (when the NC's DR errors are zero), in actuality translates and rotates. The way it moves and the things which cause it to move will be revealed when we carefully define the relative grid in terms that will enable it to be implemented.

Let us assume that each vehicle in the net, including the NC, periodically announces his own position in the relative grid. The NC's position in the relative grid, as announced by him, is defined to be perfect. This statement, turned inside out, becomes part of the definition of the relative grid. The relative grid origin and grid axis directions must be such that all times the NC's coordinates in this grid equal the coordinates he announces to the other users in his messages. This implies the following three statements: 1) error of the NC's speed sensor causes the origin of the grid to move in the direction of the NC's keel at a rate numerically equal to the error; 2) rotational drift of the NC's directional reference causes the relative navigation grid axes to rotate at the same rate; and 3) drift of the NC's directional reference at a rate $\Delta\theta_a$ causes the origin of the grid to move at a rate $R_a \Delta\theta_a$ where R_a is the range of A (the NC) from the origin (see eqn. A19). If the NC has a more sophisticated DR system, such as an inertial navigation system, then its errors will propagate into motion of the relative navigation grid in a more complicated way. The state vector definition, and its dynamics, are developed in Appendix A.

SIMULATION STUDIES

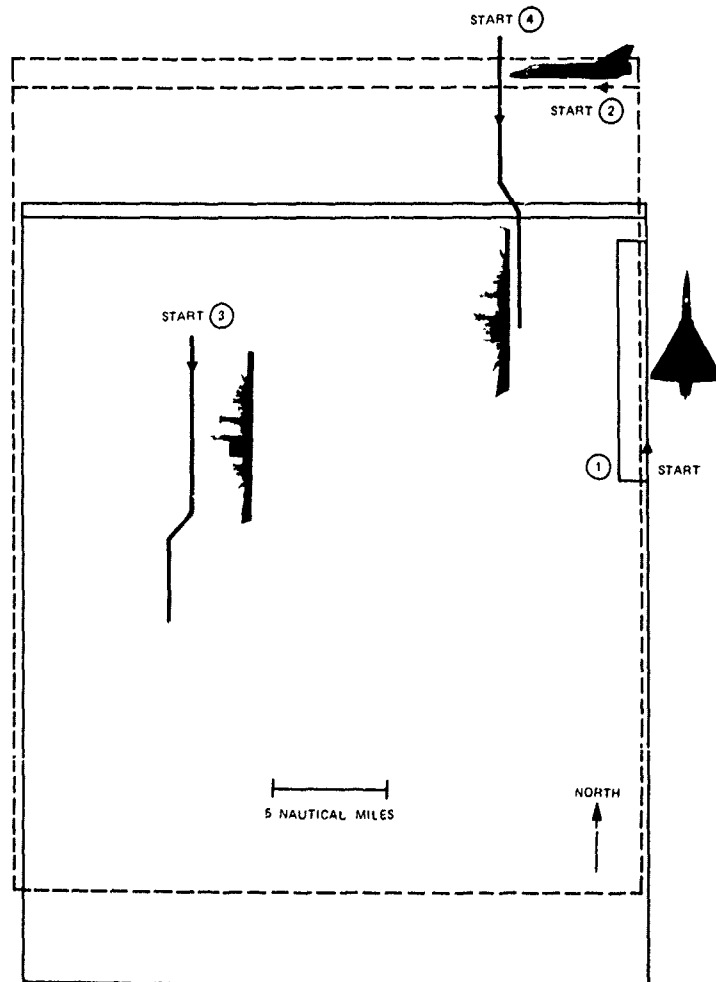
The simulation, called Simulation 4, consists of two programs, the Driver, and the Filter. The Driver sets up the vehicle geometry as a function of time (the "maneuver schedule"), and this information is provided to the Filter. The Filter contains the equations of the propagation of the covariance matrix of the extended Kalman filter. The Filter prepares selected plots of the results for visual interpretation. In particular, it provides a plot of the square root of each main-diagonal element of the covariance matrix, and three figures of merit versus measurement number in the sequence of measurements.

One particularly interesting case studied dealt with two aircraft (A and B), and two ships (C and D), with the aircraft circling the ships in a closed pattern (Fig. 1(a)). This case has good geometry because the changing directions of the lines of sight will permit triangulation; one expects to find "high" observability.

<p>Summary of Discrete Extended Kalman Filter Equations</p>	System model	$x_k = \Phi_{k-1} x_{k-1} + w_{k-1}$	$w_k \sim N(0, Q_k)$
	Measurement model	$z_k = H_k x_k + \nu_k$	$\nu_k \sim N(0, R)$
	Initial conditions	$E[x(0)] = \hat{x}_0$	$E[(x(0) - \hat{x}_0)(x(0) - \hat{x}_0)^T] = P_0$
	Other assumptions	$E[w_k \nu_j^T] = 0$	for all j, k
	State estimate extrapolation	$\hat{x}_k^- = \Phi_{k-1} \hat{x}_{k-1}^+$	
	Error covariance extrapolation	$P_k^- = \Phi_{k-1} P_{k-1}^+ \Phi_{k-1}^T + Q_{k-1}$	
	State estimate update	$\hat{x}_k^+ = \hat{x}_k^- + K_k [z_k - h_k(\hat{x}_k^-, \hat{x}_k^-)]$	
	Error covariance update	$P_k^+ = [I - K_k H_k] P_k^-$	
	Kalman gain matrix	$K_k = P_k^- H_k^T [H_k P_k^- H_k^T + R_k]^{-1}$	

Fig. 1.

(a) Group configuration.



The ships sail due south for most of the time, but each takes a short jog, one to the southwest, the other to southeast. The mission time is 30 min, during which time 31 range measurements are made. The lines of sight between the two vehicles in each measurement take on diverse directions, effectively spanning the 360° range of possibilities. This can be visualized by noting the aircraft positions according to the numbered dots in Fig. 1(b). The arrow numbered 1 points from the location of the transmitting vehicle to the location of the receiving vehicle, at the i th measurement time.

Table II presents a summary of the input data and the results. The one-sigma values of the speed (0.5 and 1.68) and heading (0.2 and 0.333) sensor accuracy were taken as generally representative of DR systems for ships and aircraft, respectively. All numerical values presented have been normalized to arbitrary units. A large value, 3, was used for the initial position standard deviation on each axis for each vehicle other than the NC, reflecting the uncertainty of each as to its location in the relative grid. Similarly, a large value, 2, was adopted for the initial angular standard deviation

Fig. 1.

(b) Measurement sequence.

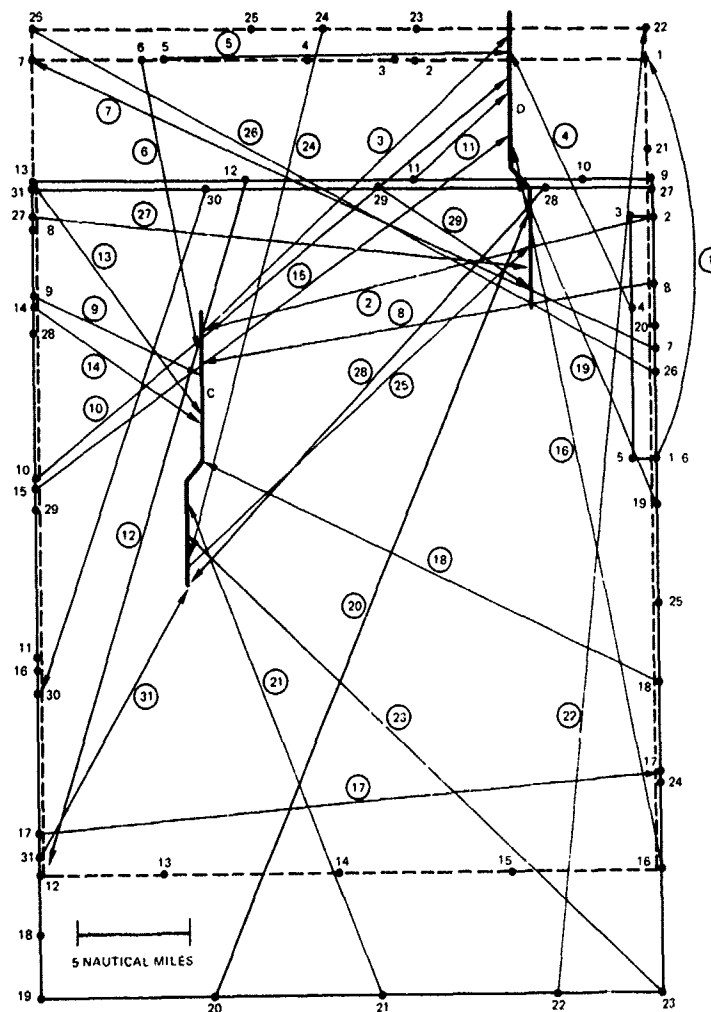


TABLE II
Summary of Input Conditions
and end results

Case	Inputs		End Results		
	Correlation Time (h)	Measurement Error (σ_R)	FOM1	FOM2	FOM3
7	3	1	0.32	0.0015	0.7
3	4	100	31.59	0.1445	68.3
4	4	100	48.51	0.2427	122.9
5	2	100	55.94	0.3003	157.7
6	1	100	65.70	0.3719	202.2

for each vehicle other than the NC, reflecting the uncertainty of each as to the orientation of the relative grid north axis compared to the indicated north of his own DR system. Since A defines the relative navigation grid, he has no initial position nor initial heading uncertainty.

Results

The results of each run are conveniently summarized by three figures of merit, FOM1, FOM2, and FOM3. Following [5], FOM1 is the rms ranging error of the network when each vehicle attempts to predict the range to each other vehicle, based on the indicated grid positions of each plus filter-estimated corrections. FOM2 is the rms angular error, when each vehicle in the net attempts to point at each other vehicle based on deriving a predicted line of sight in the relative grid, and then laying a telescope to that line of sight by measuring from the vehicle's own directional reference indicated north, corrected by his estimate of the angle to relative grid north. FOM3 is the rms cross-range miss of a hypothetical projectile fired by each vehicle at each other vehicle.

Appendixes C, D, and E derive the equations for the FOMs. Observing a plot of FOM versus measurement number is a convenient way to assess the convergence of the filter with time, that is, the ability of the users jointly to determine their position in the grid, and the angle between grid north and each's own indicated north according to his DR system. Alternatively, one can summarize the results of a run by noting the values of FOM1, 2, and 3 at the final time; if these values are acceptably small, then the filter has converged, the users have found their place in the relative grid, and the relative navigation problem has been found to be observable.

Case 7 (Table II) treats a situation with a highly accurate range measurement ($\sigma_R = 1$) and constant biases ($r = \infty$) on all DR sensors. One concludes from the results at the final time, which show very small FOMs, essentially zero that this geometry, coupled with its maneuver and measurement schedule, is highly observable. The filter is quite able to sort out all of the various individual sources of error and obtains a good estimate for each.

Having established that this is an observable geometry, the next question is, how badly do things degrade when a more realistic measurement error is introduced? Case 3 provides the answer; FOM1, the final rms ranging error, at 31.59, is proportional to the one-sigma measurement error of 100, compared with case 7. The network of users has also been able to estimate angular relationships to a satisfactory level of accuracy, as shown by FOM2 and 3 in Table II. Fig. 2 shows four computer-generated plots for this case. The first of these is the standard deviation of the error in the estimate of

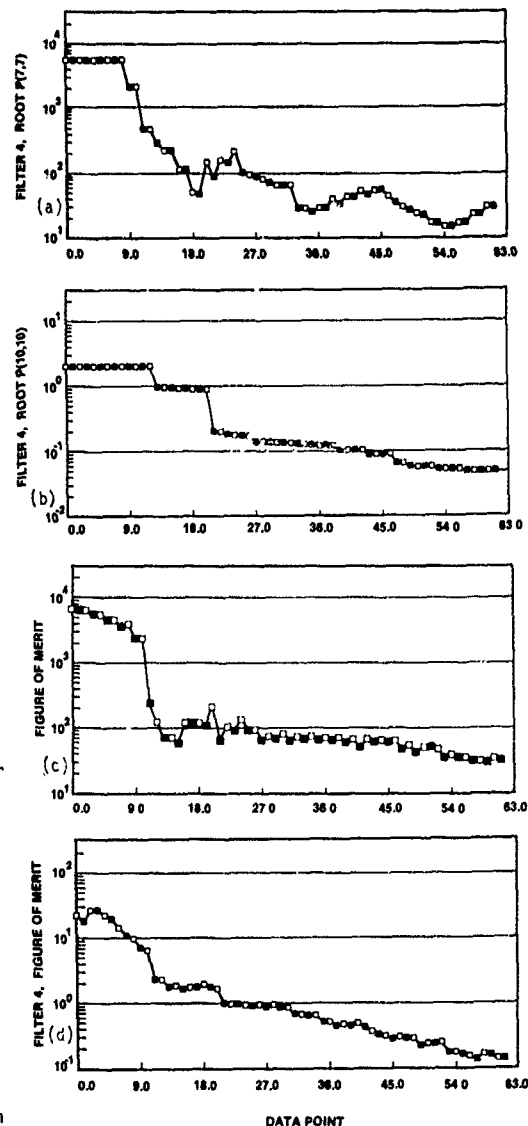


Fig. 2.

Convergence of filter, case 3. (a) Error in estimate of B's X-position, root P (7,7). (b) Error in estimate of B's directional reference with respect to relative grid north, root P (10,10). (c) RSS network ranging error, FOM1. (d) RSS network pointing error FOM2.

x_7 , vehicle B's X-position in the relative navigation grid. The second is the standard deviation of the error in the estimate of x_{10} , the angle between the Y axis of the relative grid and B's north as indicated by his directional reference. The third and fourth plots are FOM1 and FOM2. All show good convergence from initial uncertainty levels. The white and black plotted points correspond to the before- and after-measurement values, respectively.

Now suppose that the sensor errors are no longer constant, but change at a rate dependent on the correlation time τ of the associated Markov process (Appendix A). Certainly it will be more difficult for the filter to estimate changing errors than constant ones. This is borne out in the results of cases 4, 5, and 6 (Table II). Cases 3-6 form a group in which all conditions are identical, but for the correlation time. Plots of the final results for FOM1, 2, and 3 as a function of sensor correlation time appear as Fig. 3. It is seen that all errors increase as the correlation time decreases. This set of curves gives an indication of the sensitivity of the relative navigation concept to noise model parameters of the sensors. The more nearly the sensor error is constant, the better the filter can estimate its value.

Recall that these are optimal results under conditions of no mismatch. That is, the speed sensor error (for example) really does behave as a first-order Markov process with the stated correlation time and the known standard deviation, exactly as set into the program. If the actual sensor has a different response than that modeled in the filter, then the simulation result could be overly optimistic.

CONCLUSIONS

This paper has presented an optimal, centralized filter to study relative navigation observability. The effect of the errors of the navigation controller's DR sensors on the motion of the relative navigation grid has been explicitly accounted for. In a good geometry and maneuver schedule, the relative navigation positions are observable; that is, given fixed bias errors on all the DR sensors, and given perfect measurements, the rms network ranging and pointing errors are driven

essentially to zero after a suitable series of measurements. Thus, a given user is able to locate himself in the relative navigation grid and to find its orientation relative to his own directional reference. In the same geometry/maneuver schedule, and given constant DR sensor errors, but considering measurements with less than perfect accuracy, the rms network ranging accuracy is proportional to the measurement accuracy. The sensitivity of the network ranging and pointing accuracy to correlation time of the sensors has been shown for the case where the DR sensor errors are not constants, but rather changing according to Markov random processes. The results show a moderate degradation in accuracy with decreasing correlation time. The fact that the pointing accuracy, FOM2, converges, indicates that it will be feasible to exchange tracks among vehicles.

APPENDIX A. DERIVATION OF EQUATIONS OF MOTION

Discussion of the Reference Frames of Interest

Careful consideration of the relevant reference frames is essential to making explicit the motion of the relative navigation grid as a function of the errors in the sensors of the NC's dead-reckoning system. For this purpose, two frames will suffice, the relative navigation frame or grid denoted by M (for "moving") and an Earth-fixed frame denoted by F. The axes, origin, and unit vectors of F are denoted by $X_F, Y_F, Z_F, 0, i, j, k$, respectively. X_F and Y_F are nominally east and north, respectively. Z_F is up. The axes, origin, and unit vectors of the M-frame are $X, Y, Z, Q, a, \beta, \gamma$, respectively. The two frames are coincident at the start ($t = 0$) of the operation, with $Y_F = Y$ directed along A's indicated north (A being the NC) at an angle $\Delta\theta_a(0)$ from true north. The angle $\Delta\theta_a(t)$ is defined to be positive in the clockwise (CW) direction from true north to A's instantaneous indicated north, as shown in Fig. 4

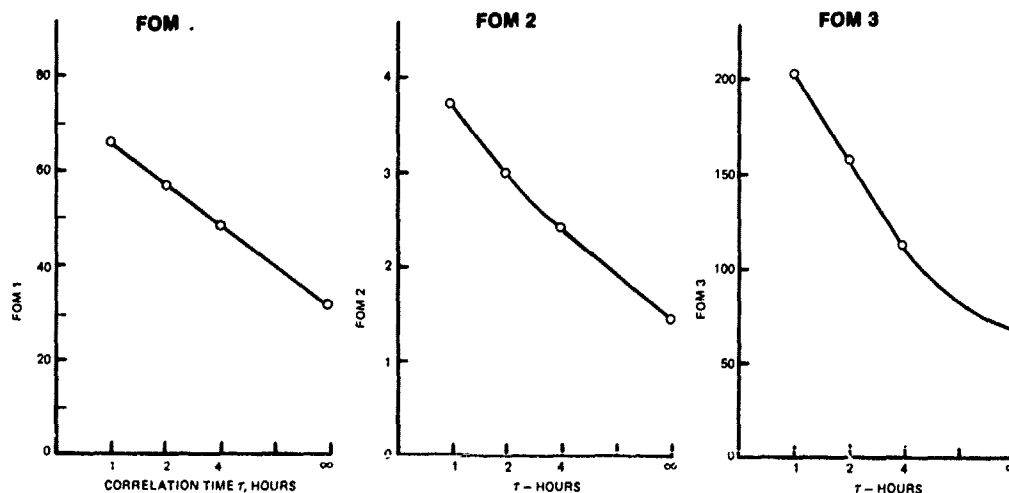
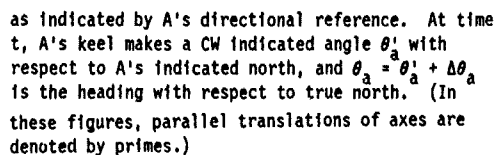


Fig. 3. Variation of convergence with correlation time.



Now look at some other user, say B. His true position coordinates in the M-frame are (x_b, y_b) .

At this instant, his directional reference generates a physical direction, "indicated north," as shown in Fig. 5. θ'_b is the indicated heading of his vehicle, where θ'_b is the CW angle from B's indicated north to the direction of B's longitudinal axis. The true heading of B is the CW angle θ_b from true north to the longitudinal axis. The correction to heading, $\Delta\theta_b$, is that angle which must be added to θ'_b to get θ_b . Therefore, $\Delta\theta_b$ is the CW angle from true north to B's indicated north. The CW angle from the relative grid Y to B's directional reference is $(\Delta\theta_b - \Delta\theta_a)$, Fig. 5. All these angular variables are functions of time; their dependence on time is frequently suppressed from the notation for convenience.

For user B to locate the relative navigation grid, he must know how his directional reference (indicated north) is pointed, relative to the indicated north of the directional reference of A; that is, he must know $(\delta\theta_b - \delta\theta_a)$. It will be the job of the Kalman filter to estimate the correction state $(\delta\epsilon_b - \delta\epsilon_a)$.

Fig. 4. Reference frames at time zero

for time $t = 0$. The Z_F and Z axes remain parallel to each other, directed upward along the true vertical for $t \geq 0$. As time progresses, A moves, and so does the M -frame, until the situation is as pictured in Fig. 5. Here Q has moved away from O by distances U and V along the $X_F Y_F$ axes. The axes of M have rotated, and Y is now a CW angle $\Delta\theta_z(t)$ away from true north. At every instant, the Y axis is defined to be parallel to the direction of north

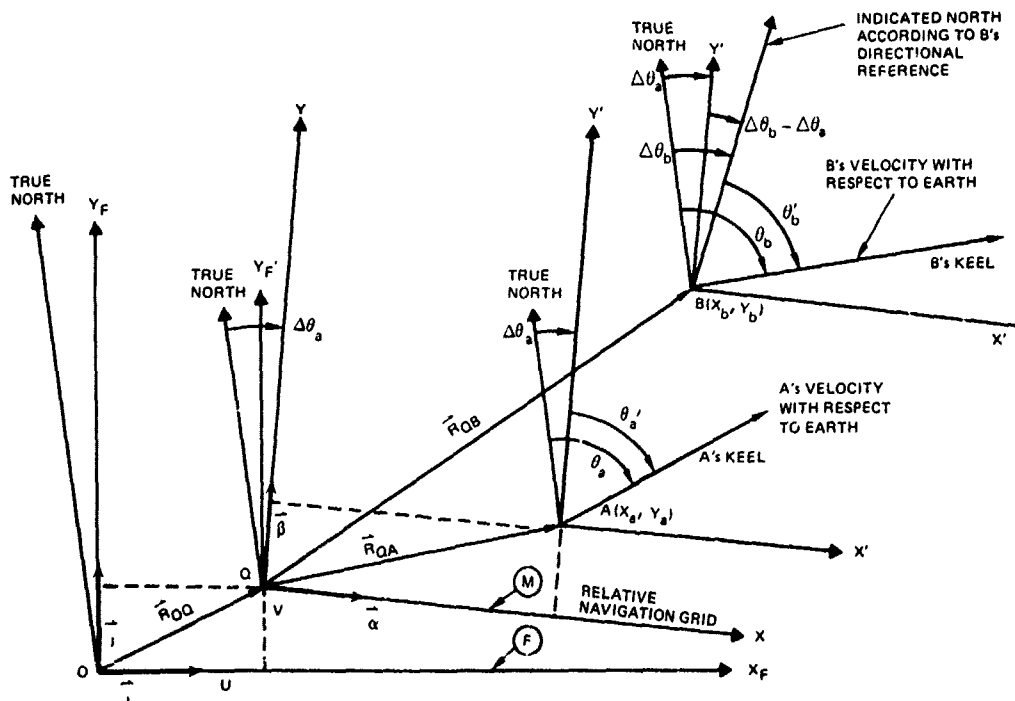


Fig. 5. Definition of angles and vectors for user B. (Symbols with arrows above them correspond to boldface symbols in the text.)

The Motion of A

Consider the motion of A. Referring to Fig. 5,

$$R_{OA} = R_{OQ} + R_{QA} \quad (A1)$$

Differentiate with respect to the F-frame and convert the last term on the right-hand side to a derivative with respect to the M-frame, using Coriolis' law, where W_{FM} is the angular velocity of the M-frame with respect to the F-frame

$$\left. \frac{dR_{OA}}{dt} \right|_F = \left. \frac{dR_{OQ}}{dt} \right|_F + \left. \frac{dR_{QA}}{dt} \right|_M + W_{FM} \times R_{QA} \quad (A2)$$

The procedure to be followed next will be to express all vectors in terms of M-frame components, and to use first-order approximations where appropriate for products involving small quantities. The set of unit vectors i, j in terms of α, β are obtained by inspection of Fig. 5:

$$i \approx \alpha + \beta \Delta\theta_a \quad j \approx -\alpha \Delta\theta_a + \beta \quad (A3)$$

where the symbol \approx implies "equal to first order."

Also, α, β in terms of i, j are

$$\alpha \approx i - j \Delta\theta_a \quad \beta \approx -i \Delta\theta_a + j \quad (A4)$$

By definition of U and V , $R_{OQ} = iU + jV$. Differentiating with respect to the F-frame, and expressing the result in terms of its M-frame components through the use of (A3),

$$\left. \frac{dR_{OQ}}{dt} \right|_F = (\alpha + \beta \Delta\theta_a) \dot{U} + (-\alpha \Delta\theta_a + \beta) \dot{V} \approx \alpha \dot{U} + \beta \dot{V} \quad (A5)$$

where we have dropped second-order terms in \dot{U}, \dot{V} , and $\Delta\theta_a$, all of which are small quantities. By definition of the coordinates of A in the M-frame,

$$R_{QA} = \alpha X_a + \beta Y_a \quad (A6)$$

Differentiating with respect to the M-frame

$$\left. \frac{dR_{QA}}{dt} \right|_M = \alpha \dot{X}_a + \beta \dot{Y}_a \quad (A7)$$

By definition, and noting that $\Delta\theta_a$ is positive when the Y axis drifts in the CW direction,

$$W_{FM} = -\gamma \Delta\theta_a \quad (A8)$$

The cross-product in (A2) is evaluated using (A6) and (A8).

$$W_{FM} \times R_{QA} = \alpha Y_a \Delta\theta_a - \beta X_a \Delta\theta_a \quad (A9)$$

The vector $dR_{QA}/dt|_F$ appearing in (A2) represents the true velocity of A with respect to the Earth.

Assuming that this vector has true magnitude S_a , and direction along A's longitudinal axis (Fig. 5), we can express this velocity in terms of its M-frame coordinates as follows:

$$\left. \frac{dR_{OA}}{dt} \right|_F = \alpha S_a \sin \theta'_a + \beta S_a \cos \theta'_a \quad (A10)$$

Inserting (A10), (A5), (A7), and (A9) into (A2), leads to the following equations, one each for the α and β components:

$$\alpha: S_a \sin \theta'_a = \dot{U} + \dot{X}_a + Y_a \Delta\dot{\theta}_a \quad (A11)$$

$$\beta: S_a \cos \theta'_a = \dot{V} + \dot{Y}_a - X_a \Delta\dot{\theta}_a \quad (A12)$$

In these equations, \dot{X}_a and \dot{Y}_a are the east and north components of A's velocity in the relative navigation grid. They are obtained as outputs of A's DR system, using the measured speed and measured heading and accepting these measurements as though they were true, according to the definition that the NC's position in the relative navigation grid is perfect. That is,

$$\dot{X}_a = S'_a \sin \theta'_a \quad \dot{Y}_a = S'_a \cos \theta'_a \quad (A13)$$

Inserting these into (A11) and (A12) gives

$$S_a \sin \theta'_a = \dot{U} + S'_a \sin \theta'_a + Y_a \Delta\dot{\theta}_a \quad (A14)$$

$$S_a \cos \theta'_a = \dot{V} + S'_a \cos \theta'_a - X_a \Delta\dot{\theta}_a \quad (A15)$$

By the definition of correction states,

$$S'_a = S_a - \Delta S_a \quad (A16)$$

Use (A16) to eliminate S'_a from (A14) and (A15) and obtain

$$0 = \dot{U} - (\sin \theta'_a) \Delta S_a + Y_a \Delta\dot{\theta}_a \quad (A17)$$

$$0 = \dot{V} - (\cos \theta'_a) \Delta S_a + X_a \Delta\dot{\theta}_a \quad (A18)$$

Now replace θ'_a by $(\theta_a - \Delta\theta_a)$, expand the trigonometric functions, and drop all second-order terms in the small quantities $\Delta\theta_a$ and ΔS_a . Then solving for \dot{U} and \dot{V} , we obtain

$$\left. \begin{aligned} \dot{U} &= (\sin \theta_a) \Delta S_a - Y_a \Delta\dot{\theta}_a \\ \dot{V} &= (\cos \theta_a) \Delta S_a - X_a \Delta\dot{\theta}_a \end{aligned} \right\} \quad (A19)$$

This pair of equations tells us how the relative grid origin is translated by the speed and heading errors of A's DR system.

Equation (A19) is part of the total set needed to define the dynamics of the microstates, according to the matrix dynamics equation $\dot{\mathbf{x}} = \mathbf{F}\mathbf{x} + \mathbf{w}$, which leads to the difference equations of the system, the first line of Table I.

The Motion of B

We analyze the motion of user B in a manner similar to that for A, taking note, however, of the differences which arise due to the preferred treatment accorded to A by the rest of the relative navigation network. The true position vector for B can be expressed in terms of its M-frame coordinates (see Fig. 5).

$$\mathbf{R}_{QB} = \alpha \mathbf{x}_b + \beta \mathbf{y}_b. \quad (\text{A20})$$

Differentiating with respect to the M-frame,

$$\left. \frac{d\mathbf{R}_{QB}}{dt} \right|_M = \alpha \dot{\mathbf{x}}_b + \beta \dot{\mathbf{y}}_b. \quad (\text{A21})$$

From Fig. 5,

$$\mathbf{R}_{OB} = \mathbf{R}_{OQ} + \mathbf{R}_{QB}.$$

Differentiating with respect to the F-frame, but then converting the last term to the M-frame with Coriolis' law:

$$\begin{aligned} \left. \frac{d\mathbf{R}_{OB}}{dt} \right|_F &= \left. \frac{d\mathbf{R}_{OQ}}{dt} \right|_F + \left. \frac{d\mathbf{R}_{QB}}{dt} \right|_F \\ &= \left. \frac{d\mathbf{R}_{OQ}}{dt} \right|_F + \left. \frac{d\mathbf{R}_{QB}}{dt} \right|_M + \mathbf{W}_{FM} \times \mathbf{R}_{QB}. \end{aligned} \quad (\text{A22})$$

Express each vector in this equation in terms of its M-frame components. The vector on the left-hand side is the true velocity of B with respect to the Earth. From (A3), this is seen to be

$$\begin{aligned} \left. \frac{d\mathbf{R}_{OB}}{dt} \right|_F &= \alpha S_b \sin \left\{ \theta'_b + \Delta\theta_b - \Delta\theta_a \right\} \\ &\quad + \beta S_b \cos \left\{ \theta'_b + \Delta\theta_b - \Delta\theta_a \right\}. \end{aligned} \quad (\text{A23})$$

Inserting (A5), (A21), (A8), (A20), and (A23) into (A22), separating by α , β components, and solving for $\dot{\mathbf{x}}_b$, $\dot{\mathbf{y}}_b$, the true velocity components of B in the relative grid,

$$\dot{\mathbf{x}}_b = S_b \sin \left\{ \theta'_b + \Delta\theta_b - \Delta\theta_a \right\} - \dot{U} - Y_b \Delta\dot{\theta}_a \quad (\text{A24})$$

$$\dot{\mathbf{y}}_b = S_b \cos \left\{ \theta'_b + \Delta\theta_b - \Delta\theta_a \right\} - \dot{V} + X_b \Delta\dot{\theta}_a \quad (\text{A25})$$

B's indicated components of velocity in the relative grid ($\dot{\mathbf{x}}'_b$, $\dot{\mathbf{y}}'_b$) are obtained from his DR system:

$$\dot{\mathbf{x}}'_b = S'_b \sin \theta'_b \quad \dot{\mathbf{y}}'_b = S'_b \cos \theta'_b. \quad (\text{A26})$$

Using the convention for corrections,

$$\Delta\dot{\mathbf{x}}'_b = \dot{\mathbf{x}}'_b - \dot{\mathbf{x}}_b; \quad \Delta\dot{\mathbf{y}}'_b = \dot{\mathbf{y}}'_b - \dot{\mathbf{y}}_b. \quad (\text{A27})$$

Substituting (A24), (A25), and (A26) into (A27), along with the definition $S'_b = S'_b + \Delta S'_b$, and expanding the trigonometric functions to first order in small quantities, we obtain the differential equations for B's position-correction states,

$$\begin{aligned} \Delta\dot{\mathbf{x}}'_b &\approx \left\{ \sin \theta'_b \right\} \Delta S_b + \left\{ S_b \cos \theta'_b \right\} \left\{ \Delta\theta_b - \Delta\theta_a \right\} \\ &\quad - \dot{U} - Y_b \Delta\dot{\theta}_a \end{aligned} \quad (\text{A28})$$

$$\begin{aligned} \Delta\dot{\mathbf{y}}'_b &\approx \left\{ \cos \theta'_b \right\} \Delta S_b - \left\{ S_b \sin \theta'_b \right\} \left\{ \Delta\theta_b - \Delta\theta_a \right\} \\ &\quad - \dot{V} + X_b \Delta\dot{\theta}_a. \end{aligned} \quad (\text{A29})$$

Similar equations hold for the states associated with vehicles C and D, obtained by replacing subscript b with c and d in turn.

The State Vector

The state vector \mathbf{x} of the Kalman filter can now be defined. Vehicle B has four states associated with it, the correction to east (ΔX_b) and north (ΔY_b) position in the relative grid, the correction to speed (ΔS_b), and the CW angle ($\Delta\theta_b - \Delta\theta_a$) from relative navigation grid north to B's directional reference. These four states are $x_7 - x_{10}$, respectively. Vehicle C has four similar states ($x_{11} - x_{14}$), and D has four ($x_{15} - x_{18}$). A is treated differently because of his special role as NC. There are no corrections to his relative navigation position. We do try to estimate $\Delta\theta_a$ for him; however, in view of the fact that there is nearly nothing in the system that makes true north directly observable (except for the weak tendency of all four directional references to distribute their errors about true north over time), we will not expect estimates for $\Delta\theta_a$ to be particularly good. The correction states associated with A are his speed correction (ΔS_a), his heading correction ($\Delta\theta_a$) and its derivative ($\Delta\dot{\theta}_a$), the Cartesian components U and V of the displacement of the origin of the relative grid from its initial position, and the initial value $\Delta\theta_a(0)$. (This last is a remnant of a previous formulation of the problem and could have been omitted.) The states, 18 in all, are defined in Table III.

State Vector Dynamics

A first-order differential equation must be found for each component of the state vector. Starting with $x_1 \equiv \Delta S_a$, a first-order Markov process with correlation time $\tau_{sa} \equiv \tau_1$ is used to model A's speed-sensor correction. The same kind

TABLE III
The State Vector of the Kalman Filter

Definitions
$x_1 = \Delta S_a$
$x_2 = \Delta \theta_a$
$x_3 = \Delta \theta_a^b$
$x_4 = U$
$x_5 = V$
$x_6 = \Delta \theta_a(0)$
$x_7 = \Delta X_b$
$x_8 = \Delta Y_b$
$x_9 = \Delta S_b$
$x_{10} = \Delta \theta_b - \Delta \theta_a$
$x_{11} = \Delta X_c$
$x_{12} = \Delta Y_c$
$x_{13} = \Delta S_c$
$x_{14} = \Delta \theta_c - \Delta \theta_a$
$x_{15} = \Delta X_d$
$x_{16} = \Delta Y_d$
$x_{17} = \Delta S_d$
$x_{18} = \Delta \theta_d - \Delta \theta_a$

of model was used for the speed sensor of vehicles B, C, and D, and for the heading sensor of vehicles B, C, and D. For convenience, a second-order Markov process was adopted for the heading sensor of A. Differential equations have been obtained for U, V, in (A19), requiring only the change of notation to state variables. Also, differential equations have been obtained for $\Delta \theta_b$ and $\Delta \theta_c$ in (A28) and (A29). Equations of the identical form hold for vehicles C and D. These equations in the continuous state-variable form $\dot{x} = Fx + w$, can be converted to the matrix difference equation $x_k = \phi_{k-1} x_{k-1} + w_{k-1}$, as needed for the system model of Table I, by following the procedures of [6, p.77], using

$$\phi_k \approx I + F(t_k) \Delta t, \quad Q_k \approx Q(t_k) \Delta t. \quad (A30)$$

APPENDIX B. THE H MATRIX

The true range between vehicles i and j is

$$R = \left[(x_j + \Delta X_j - x_i - \Delta X_i)^2 + (y_j + \Delta Y_j - y_i - \Delta Y_i)^2 \right]^{1/2}. \quad (B1)$$

Evaluate at $\Delta X = \Delta Y = 0$ to get indicated range R' :

$$R' = \left[(x_j - x_i)^2 + (y_j - y_i)^2 \right]^{1/2}. \quad (B2)$$

Expanding R in a Taylor series about $\Delta X = \Delta Y = 0$ and dropping higher order terms,

$$R = R' + a_{ij} (\Delta X_j - \Delta X_i) + \beta_{ij} (\Delta Y_j - \Delta Y_i) \quad (B3)$$

where we define a_{ij} and β_{ij} as follows:

$$a_{ij} = \frac{x_j - x_i}{R_{j1}}; \quad \beta_{ij} = \frac{y_j - y_i}{R_{j1}}. \quad (B4)$$

The state vector x has elements for each incremental variable in (B3); the coefficient of the ith state variable is H_i , by definition of (5). Hence, there are four nonzero elements in H when B, C, or D range on each other. There are only two nonzero elements in H when A is involved in the measurement, because, by definition of his role as NC, there are no Δ positions for A.

APPENDIX C. DERIVATION OF FOM1

FOM1 is the rms network ranging error, a measure of the standard deviation of the error when the member of one net predicts the range to another; that is,

$$FOM1 = \left\{ \frac{1}{6} \left[E[R_{ab}^2] + E[R_{ac}^2] + E[R_{ad}^2] + E[R_{bc}^2] + E[R_{bd}^2] + E[R_{cd}^2] \right] \right\}^{1/2} \quad (C1)$$

where \hat{R}_{jk} is the error in estimate \hat{R}_{jk} of range R_{jk} from vehicle j to vehicle k and $E[R_{jk}^2]$ is the variance of the error \hat{R}_{jk} . Since $\hat{R}_{jk} = \hat{R}_{kj}$, it is only necessary to consider the six possibilities given in (C1). The true range in the M-frame is

$$R_{jk} = \left[(x_k - x_j)^2 + (y_k - y_j)^2 \right]^{1/2} = \left[(x_k' + \Delta \hat{x}_k + \Delta \tilde{x}_k - x_j' - \Delta \hat{x}_j - \Delta \tilde{x}_j)^2 + (y_k' + \Delta \hat{y}_k + \Delta \tilde{y}_k - y_j' - \Delta \hat{y}_j - \Delta \tilde{y}_j)^2 \right]^{1/2}. \quad (C2)$$

The estimated range in the M-frame, as made by either vehicle j or k, is

$$\hat{R}_{jk} = \left[(x_k' + \Delta \hat{x}_k - x_j' - \Delta \hat{x}_j)^2 + (y_k' + \Delta \hat{y}_k - y_j' - \Delta \hat{y}_j)^2 \right]^{1/2}. \quad (C3)$$

The error in the estimate is obtained from the two preceding equations, using a Taylor series, and dropping higher order terms in the small quantities $\Delta \tilde{x}$:

$$\hat{R}_{jk} = \hat{R}_{jk} - R_{jk} = a_{jk} (\Delta \tilde{x}_k - \Delta \tilde{x}_j) + \beta_{jk} (\Delta \tilde{y}_k - \Delta \tilde{y}_j) \quad (C4)$$

where a_{jk} and β_{jk} are defined similarly to (B4), but dropping primes. The variance of the error in

the range from b to c, for example, is obtained from (C4) as follows:

$$\begin{aligned} E\{\hat{R}_{bc}^2\} &= E\{[a_{bc}(\tilde{x}_{11} - \tilde{x}_7) + \beta_{bc}(\tilde{x}_{12} - \tilde{x}_8)]^2\} \\ &= a_{bc}^2 [P_{11,11} + P_{77} - 2P_{7,11}] \\ &\quad + \beta_{bc}^2 [P_{12,12} + P_{88} - 2P_{8,12}] \\ &\quad + 2a_{bc}\beta_{bc}[P_{11,12} - P_{8,11} - P_{7,12} + P_{78}] \end{aligned} \quad (C5)$$

where we have used the definition of the covariance elements:

$$P_{jk} = E[\tilde{x}_j \tilde{x}_k]. \quad (C6)$$

APPENDIX D. DERIVATION OF FOM2

FOM2 is the rms angular pointing error when each vehicle in the net tries to point a telescope at every other vehicle on the basis of knowing the indicated position coordinates of each vehicle, together with the estimated state-vector \hat{x} . The pointing will be in error because of errors in the estimate. In addition, one must consider what is meant by "pointing." Here we mean that, if B is to point at C, for example, that B will train a telescope on his deck (assumed to be level) through an angle from B's north, as indicated by B's directional reference (gyrocompass), corrected by the estimated correction state \hat{x}_{10} , to the azimuth which B computes as the azimuth of the line of sight to C. Then the error is the difference between where the telescope line is, and the true line of sight from B to C, i.e., the error in the physical pointing angle.

In this case, because the directional reference errors are different for each vehicle, the angular error in B pointing at C is generally different from the error in C pointing at B. Thus, we have to consider all possible combinations of one vehicle pointing at another. Let

$$\text{VOPPE}_{ij} \equiv \text{variance of the physical pointing error from vehicle } i \text{ to vehicle } j, \quad i, j = 1, 2, 3, 4. \quad (D1)$$

Then FOM2, defined as

$$\text{FOM2} = \left[\frac{1}{12} \sum_{i=1}^4 \sum_{j=1, j \neq i}^4 \text{VOPPE}_{ij} \right]^{1/2} \quad (D2)$$

is the rms pointing error of the network of users.

Define ψ_{ij} = the azimuth angle of the line of sight (LOS) from vehicle i to vehicle j, measured clockwise from the north axis of the M-frame to the LOS, as in Fig. 6. Suppose B wants to point to C. The correct pointing angle S_{bc} from B's directional reference to the LOS is

$$S_{bc} = \psi_{bc} - (\Delta\theta_b - \Delta\theta_a) = \psi_{bc} - \hat{x}_{10} \quad (D3)$$

as shown in Fig. 6. The estimated value \hat{S}_{bc} is

$$\hat{S}_{bc} = \hat{\psi}_{bc} - \hat{x}_{10} \quad (D4)$$

and the error in the estimate is

$$\begin{aligned} \hat{S}_{bc} &= \hat{S}_{bc} - S_{bc} = \hat{\psi}_{bc} - \psi_{bc} - \hat{x}_{10} + \hat{x}_{10} \\ \hat{S}_{bc} &= \hat{\psi}_{bc} - \psi_{bc} - \tilde{x}_{10}. \end{aligned} \quad (D5)$$

The true value of the azimuth angle is

$$\begin{aligned} \psi_{bc} &= \tan^{-1} \left[\frac{X_c - X_b}{Y_c - Y_b} \right] \\ &= \tan^{-1} \left[\frac{X'_c + \Delta\hat{X}_c - \Delta\tilde{X}_c - X'_b - \Delta\hat{X}_b + \Delta\tilde{X}_b}{Y'_c + \Delta\hat{Y}_c - \Delta\tilde{Y}_c - Y'_b - \Delta\hat{Y}_b + \Delta\tilde{Y}_b} \right] \end{aligned} \quad (D6)$$

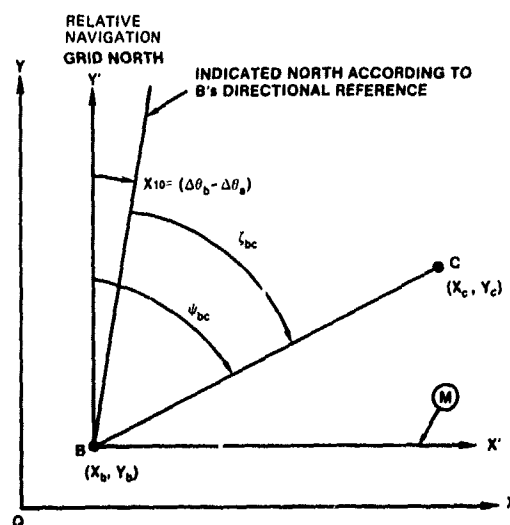


Fig. 6. Telescope-pointing angles.

The estimated value is

$$\hat{\psi} = \tan^{-1} \left[\frac{x'_c + \Delta \hat{x}_c - x'_b - \Delta \hat{x}_b}{y'_c + \Delta \hat{y}_c - y'_b - \Delta \hat{y}_b} \right]. \quad (D7)$$

Substituting (D6) and (D7) into (D5), expanding to first order in the small quantities \tilde{x} , and replacing the ΔX 's and ΔY 's by their state-vector counterparts, we obtain

$$\tilde{s}_{bc} = -\tilde{x}_{10} + \frac{\beta_{bc}}{R_{bc}} (\tilde{x}_{11} - \tilde{x}_7) - \frac{a_{bc}}{R_{bc}} (\tilde{x}_{12} - \tilde{x}_8). \quad (D8)$$

The variance of this error is

$$\begin{aligned} \text{VOPPE}_{23} = & P_{10,10} + \left(\frac{\beta_{bc}}{R_{bc}} \right)^2 (P_{11,11} - P_{7,11} + P_{77}) \\ & + \left(\frac{a_{bc}}{R_{bc}} \right)^2 (P_{12,12} - 2P_{8,12} + P_{88}) \\ & + \frac{2\beta_{bc}}{R_{bc}} (-P_{10,11} + P_{7,10}) \\ & + \frac{2a_{bc}}{R_{bc}} (P_{10,12} - P_{8,10}) \\ & - \frac{2a_{bc}\beta_{bc}}{R_{bc}^2} (P_{11,12} - P_{8,11} - P_{7,12} + P_{78}). \end{aligned} \quad (D9)$$

The other VOPPEs are similarly constructed, except that when the NC is involved, because of the perfect knowledge of his location in the grid, $\Delta \hat{x}_a = \Delta \hat{y}_a = 0$. Also, because of his perfect knowledge of grid orientation, when A does the pointing, the error in the estimate of ψ_{ak} is the total error, so somewhat simpler expressions than (D9) are obtained.

APPENDIX E. DERIVATION OF FOM3

FOM3 is the rms linear crossrange miss when each vehicle in the net tries to fire an imaginary projectile at each other member. For one pair of vehicles i and j , the linear miss when i fires at j is the product of the angular pointing error (ij) times the range R_{ij} . Therefore, the variance of this error, to first order, is R_{ij}^2 times VOPPE_{ij} , and the rms crossrange error of the network is

$$\text{FOM3} = \left[\frac{1}{12} \sum_{\substack{j=1 \\ j \neq i}}^4 \sum_{i=1}^4 R_{ij}^2 \text{VOPPE}_{ij} \right]^{1/2} \quad (E1)$$

REFERENCES

- [1] Schneider, A.M. (1982) Observability of relative navigation using range-only measurements. In Proceedings of IEEE Position Location and Navigation Symposium (PLANS '82) (Atlantic City, N.J., Dec. 1982).
- [2] Schneider, A.M. Studies in observability II. NOSC Report, Naval Ocean Systems Center, San Diego, Calif. To be published.
- [3] Gobbini, G.F. (1981) Relative navigation by means of passive rangings. Ph.D. dissertation, Department of Aeronautics and Astronautics, Massachusetts Institute of Technology, Cambridge, Mass., p. 60, 1981.
- [4] Widnall, W.S., and Gobbini, G.F. (1982) Community relative navigation based on measurement sharing. Presented at the IEEE Position, Location, and Navigation Symposium (PLANS 82) (Atlantic City, N.J., Dec. 1982).
- [5] Rome, H.J., and Stambaugh, J.S. (1977) Evaluation of the accuracy of community relative navigation organization concepts. Navigation, 24, 2 (Summer 1977), 168-180.
- [6] Gelb, A. (Ed.) (1974) Applied Optimal Estimation. Cambridge, Mass.: M.I.T. Press, 1974.
- [7] Brown, R.G. (1972) Integrated navigation systems and Kalman filtering: a perspective. Navigation, 19, 4 (Winter 1972-1973), 355-362.

ACKNOWLEDGEMENTS

The author would like to thank R. Akita for direction and support, G. Vegter for many helpful discussions, and K. Schemensky for programming.

INTEGRATED STRAPDOWN AVIONICS FOR PRECISION GUIDED VEHICLES

by

Jack Richman, David Haessig, Jr., and Bernard Friedland
The Singer Company, Kearfott Division
1150 McBride Avenue, Little Falls, NJ 07424
United States

ABSTRACT

Conventional avionic configurations for precision guided vehicles are often unnecessarily costly and inefficient because of built-in (but unused) redundancy in instrumentation attributed to the present day independent systems design approach. Described in this paper is an integrated design approach using strapdown avionic components that has the potential for lowering cost, increasing reliability, and improving overall performance as a result of using fewer and less costly instruments in an efficient manner.

INTRODUCTION

The avionics associated with present day autonomously guided vehicles (missiles, glide bombs, drones, etc.) have conventionally been configured by design engineers as being comprised of 3 separate and independent systems; namely, navigation, guidance (midcourse and/or terminal) and autopilot.

The navigation system, whose purpose it is to provide position, velocity and attitude of the vehicle with respect to some reference coordinate frame is conventionally configured as an inertial system equipped with relatively high accuracy components (gyros and accelerometers) in either a gimbaled or strapdown mode.

The guidance system, whose purpose it is to generate midcourse and/or terminal steering commands is, in many applications, configured as a system having an inertially stabilized seeker that directly measures some combination of range, angles and angular rates between the vehicle and an aim point in some fixed guidance coordinate frame.

The autopilot, whose purpose it is to match the commanded acceleration of the guidance system by issuing its own commands to appropriate aerodynamic and/or thrust controls of the vehicle, is usually configured as a system equipped with low accuracy inertial components (gyros and accelerometers).

Clearly, the above conventional avionic configuration of a precision guided vehicle lends itself to duplication and redundancy of components (gyros, accelerometers and gimbals). Yet, in spite of this obvious redundancy, no attempt is made in conventional designs to combine the multiplicity of output data from these individual systems in some efficient manner; and furthermore, no provision is made to channel the redundant data from one system to another in case of component failure. This obvious inefficiency in the conventional avionic design process has its roots in the fact that design of the navigation, guidance and autopilot systems has historically been performed by three separate design groups, each having different disciplines and each meeting the required specifications of their own system with a self contained design.

With the avionics portion of precision guided vehicles representing a significant portion of the cost, an obvious approach to cost reduction is through the use of integrated avionics: a common set of components shared by all systems. In addition to having lower cost and higher reliability (as a result of using fewer and less costly instruments in an efficient manner), in many cases superior performance can be achieved with this integrated strapdown design approach. Described in this paper is an approach for designing such an integrated system using strapdown instruments and some unusual examples of their use.

THE INTEGRATED STRAPDOWN DESIGN

We limit our discussion to a vehicle having a self contained avionic system, although the concept is readily extended to the case in which data from external aids (e.g. GPS) is available.

In our integrated avionics design approach for a precision guided vehicle, the following basic components are considered:

- A single set of strapdown gyros and accelerometers to be shared in all functions of navigation, guidance and control.
- A strapdown midcourse and/or terminal seeker.
- A computer that includes Kalman mixer-filter computations for combining the avionics data in an optimum manner and the steering and autopilot computations.

The general structure of such an integrated strapdown avionics systems for steering the vehicle to an aimpoint is shown in Figure 1. We will refer to the aimpoint of the vehicle as the "target", and indeed the aimpoint could represent a physical target to be intercepted by a guided weapon, or it could be a precise point to which a drone is to be steered (relative to the sighting point used by the seeker). In the guided vehicle applications presented in this paper, the seeker sighting point was taken to be the target.

Unlike conventional navigation, the guided vehicle performs its navigation with respect to an imprecisely known, target location which may not even be stationary. For simplicity, the discussion of the system's operation is limited to the case in which the target is nonmaneuvering (i.e., a nonaccelerating target). Although the integrated strapdown concept is also applicable to a maneuvering target, the problem becomes more complex and requires an algorithm for estimating the target's acceleration [1], [2].

In describing the operation of the overall system shown in Fig. 1, it is convenient to first trace the path of the inertial subsystem through the navigation, guidance, and autopilot functions as if the inertial subsystem operated independently, and then to show the mutual aiding that exists between the seeker and inertial subsystems. The navigation computation starts with an initial estimate of the inertial position and velocity of the vehicle with respect to its target, and knowledge of the inertial attitude of the vehicle. The inertial sensors (strapdown gyros and accelerometers) perform the continuous updating of the position, velocity and attitude of the vehicle. This state of the vehicle with respect to the target can readily be transformed into appropriate guidance states to be used as input to the guidance (or steering) law. For example, if the guidance state is the inertial line-of-sight rate vector λ between the missile and its target [as would be the case for simple proportional-navigation (ProNav) guidance], the transformation between the navigation state and guidance state is

$$\dot{\lambda} = (R \times V) / |R|^2 \quad (1)$$

The output of the guidance law is usually a commanded acceleration a_c which the autopilot attempts to achieve by issuing commands to its appropriate controllers, which in the case of an aerodynamically controlled vehicle are commanded aerodynamic surface deflections δ_c . (Note that the autopilot also uses the same inertial sensors as those used for navigation.)

From the above description of the guidance system, it is clear that if the navigation computation started with perfect estimates of the initial state of the vehicle with respect to its target and if the inertial sensors were error-free, the vehicle would not require a seeker. The seeker is needed only because the inertial system has errors associated with it. The primary function of the seeker in this integrated configuration is to provide corrections to the inertial navigation/guidance system. (For the case in which the guided vehicle is directed toward a maneuvering target, the role of the seeker, in addition to aiding the inertial navigation system, would be to provide estimates of target acceleration to be incorporated into the navigation computation block shown in Fig. 1.) The lower portion of the block diagram of Fig. 1 shows the role of the seeker and its interaction with the inertial system. The seeker system is shown as consisting of a strapdown seeker "head" plus data processing. The head provides the basic strapdown measurements of a target with respect to the vehicle's (body) coordinate frame. The data processing block combines the strapdown measurements of the head with the vehicle attitude to produce seeker measurements with respect to an inertial reference frame. For example, if the seeker head measured discrete line-of-sight angles with respect to the vehicle's coordinate frame, the output of the seeker and the data processor would be line-of-sight angles with respect to an inertial coordinate frame. In particular, if the strapdown seeker measurements in the vertical and horizontal planes of the vehicle (body line-of-sight angles) are λ_{VB} and λ_{HB} , respectively, and the transformation from the inertial coordinate frame to the body coordinate frame (as obtained from processing the strapdown gyro data) is denoted by the transformation matrix

$$T_{BI} = \begin{bmatrix} l_1 & l_2 & l_3 \\ m_1 & m_2 & m_3 \\ n_1 & n_2 & n_3 \end{bmatrix}$$

the inertial line-of-sight angles are computed in accordance with

$$\lambda_{VI} = \tan^{-1} \left\{ \frac{l_3 + m_3 \tan \lambda_{HB} + n_3 \tan \lambda_{VB}}{l_1 + m_1 \tan \lambda_{HB} + n_1 \tan \lambda_{VB}} \right\}$$

$$\lambda_{HI} = \tan^{-1} \left\{ \frac{l_2 + m_2 \tan \lambda_{HB} + n_2 \tan \lambda_{VB}}{l_1 + m_1 \tan \lambda_{HB} + n_1 \tan \lambda_{VB}} \right\}$$

An independent estimate of the seeker output can always be obtained from the inertial navigation system. For example, if the seeker-system measurements consist of the vertical and horizontal (i.e. elevation and azimuth) angles of the target with respect to an inertial coordinate frame, the estimates of these measurements obtained from the inertial navigation system are given by

$$\hat{\lambda}_{VI} = \tan^{-1} \left\{ \hat{z}_I (\hat{x}_I^2 + \hat{y}_I^2)^{-1/2} \right\}$$

$$\hat{\lambda}_{HI} = \tan^{-1} \left\{ \hat{y}_I (\hat{x}_I^2 + \hat{y}_I^2)^{-1/2} \right\}$$

where \hat{x}_I , \hat{y}_I , and \hat{z}_I are the estimated coordinates of the vehicle with respect to the target as supplied by the navigation system. The difference, or residual, between the actual seeker measurements and the independent estimates is the input to the Kalman mixer/filter, the outputs of which are corrections to the navigation state as well as corrections to seeker systematic errors that may have been modeled and included in the Kalman filter.

The most significant feature associated with the operation of the integrated strapdown system is that the inertial navigator and seeker work together in providing the input to the guidance law. This mutual aiding of the systems can be described from two different points of view. One interpretation is that the inertial navigator is the primary system that provides the input to the guidance law and the function of the strapdown seeker is to provide corrections to the inertial system. Another interpretation (from the point of view of the seeker designer) is that the seeker is the primary instrument that provides the input to the guidance system and the function of the inertial system is to smooth the data between the discrete seeker measurements. Regardless of the interpretation, the two systems operate in an optimum integrated manner.

Early investigators who looked at the possibility of using strapdown seekers for vehicle guidance came to the conclusion that strapdown seekers cannot be used in guided weapon systems. Indeed, these early investigators were correct, in the sense that attempting to obtain useful guidance data from an unaided strapdown seeker will almost always be doomed to failure. The seeker random errors plus small systematic errors combined with high vehicle

rotational rates have a tendency to produce large errors in the derived guidance data that must be extracted from the measured strapdown seeker. It is the integration with the strapdown inertial sensors that permits the strapdown seeker to have a useful role in the overall guidance system.

Another significant advantage of the integrated strapdown approach over the independent systems approach occurs when the seeker is no longer capable of supplying useful data as a result of either losing track of the target or due to poor vehicle-to-target geometry (e.g., when the seeker becomes "blind" as a result of the target image filling the field of view of the seeker). In the conventional configuration (which would probably use an inertially stabilized seeker) all guidance commands would remain constant, based on the last seeker measurement (i.e., zero-order hold). On the other hand, with an integrated system the guidance commands are based on the inertial navigation system, which continues to operate even after the seeker stops providing data (and can also interpolate between missing data points enroute). In fact, at the point when the seeker usually stops providing useful data, the inertial system often has been quite accurately "calibrated" by the earlier seeker data so that it is capable of providing accurate guidance signals to the very end of flight. This feature is illustrated in the strapdown configurations described below.

ILLUSTRATIVE EXAMPLES

An analytical investigation was conducted to evaluate the performance that could be achieved for a number of existing guided vehicles if equipped with variously configured strapdown systems in the integrated mode previously described. The vehicles considered for this study varied from the sluggish skid-to-turn glide bombs to high-performance bank-to-turn vehicles. In all instances, we were able to demonstrate via six-degree-of-freedom simulation that relatively low-cost strapdown systems, when properly integrated are capable of accurately guiding the vehicle to its target [3]. One of the most unconventional strapdown configurations analyzed, which clearly illustrates the benefit derived from efficiently integrating the inertial and seeker components, is an inertial strapdown system aided by a synthetic aperture radar (SAR) seeker for use in an air-to-surface vehicle. A SAR is a strapdown imaging radar seeker that is capable of providing discrete measurements of the cone angle γ between the vehicle velocity vector and target (see Fig. 2) when the cone angle is above a threshold of approximately 15 degrees. The system is also capable of quite accurately measuring the distance r to the target and very crudely measuring the base angle β .

Within the framework of combining the SAR measurements with inertial navigation data, the three SAR measurements are related to the six inertial states (x, y, z, V_x, V_y, V_z) used in the Kalman filter by

$$r = (x^2 + y^2 + z^2)^{1/2}$$

$$\gamma = \cos^{-1}(\mathbf{R} \cdot \mathbf{V} / |\mathbf{R}| |\mathbf{V}|)$$

$$\beta = \cos^{-1}(\mathbf{k}_1 \cdot \mathbf{k}_2)$$

where

$$\mathbf{k}_1 = \mathbf{R} \times \mathbf{V} / |\mathbf{R} \times \mathbf{V}|$$

$$\mathbf{k}_2 = [-V_y^2(V_x^2 + V_y^2)^{-1/2}, V_x(V_x^2 + V_y^2)^{-1/2}, 0]^T$$

A ProNav guidance law was assumed in which the steering commands are proportional to the inertial line-of-sight rate. The line-of-sight rates are obtained from the estimated inertial states using Eq. (1); the resulting commanded vehicle accelerations are computed as

$$\mathbf{a}_c = -k |\mathbf{V}| \dot{\lambda}$$

A skid-to-turn autopilot was used to achieve these accelerations through a control command having the following structure:

$$\delta_c = c_1 \mathbf{a}_c + c_2 \mathbf{A} + c_3 \dot{\theta}$$

where \mathbf{A} and $\dot{\theta}$ represent accelerometer and rate gyro output, respectively.

The major problem associated with SAR-type seeker is that shortly after the vehicle starts to steer toward the target, the cone angle falls below its imaging threshold angle and ceases to provide useful data. Because of this nonlinearity associated with the observation process, one cannot apply the separation principle to the design of the guidance system. That is, one cannot assume that a separately designed optimum estimator can be linked in tandem to a separately designed optimum steering law to yield an optimum guidance system. Most steering laws would in a very short time, tend to direct the vehicle toward the target and thereby cause the SAR system to become ineffective. The errors associated with the inertial navigation system at this early stage are too large for the inertial system to provide guidance commands without the aid of the seeker. We have learned, however, that by shaping the trajectory in a manner such that the vehicle does not completely steer toward the target for approximately one-half of its flight (allowing the cone angle to remain above its 15-degree threshold), the SAR data during this portion of flight can be processed to "calibrate" the inertial system well enough for it to accurately guide the vehicle toward the target in a pure inertial mode for the remainder of the flight. Clearly, this efficient use of the above strapdown sensors can be used only in an integrated mode with a mixer/filter to combine the data in an optimum manner.

A typical simulated trajectory is shown in Fig. 3. The vehicle has an initial altitude of 1250 ft and an initial velocity of 750 ft/sec. The target is 8000 ft down range and 4000 ft cross track. In the figure, the dots appear at 0.2-sec intervals and the vehicle profiles appear at 2.0-sec intervals (the total time of flight shown in 15.7 sec). The trajectory shaping was a two-step process -- for a portion of the flight the vehicle was directed to steer to a phantom target located at the same altitude of the vehicle and 20 degrees astride the instantaneous direction of the target. (A 20-degree reference was used to guarantee the 15-degree SAR cone angle threshold.) At some appropriate point along the trajectory, the phantom target location is changed to a position directly above the actual

target but at the vehicle altitude. This point is denoted in the figure by the arrow R_1 . A short time later (denoted in the figure by the arrow R_2) the phantom target switches to a position that is coincident with the actual target, which causes the vehicle to dive down onto the target.

The ability of the Kalman filter to improve the estimates of the navigation states used in the guidance laws is illustrated in Fig. 4. Shown in the figure is the decrease in altitude error (actual and standard deviation) from an initial uncertainty of 200 ft to less than 20 ft by the midpoint of the flight. The other navigation states showed similar improvement during the flight.

Terminal miss distance accuracy was evaluated by performing Monte Carlo simulations to several stationary target locations. The circular error probability (CEP) was less than 2 m.

Although this method of trajectory shaping achieved quite acceptable accuracy, it is admittedly ad hoc. A more systematic method of shaping the trajectory to minimize terminal miss distance in critical scenarios merits further study.

A second example illustrating an unusual benefit that might be obtained with an integrated strapdown design is that of a high performance guided missile engaged in an air-to-air encounter. The missile is an aircraft-launched, high thrust, aerodynamically steered missile having a bank-to-turn autopilot capable of achieving a 50g maneuver. The missile's seeker consists of a "forward-looking" radar (having a restricted field of view) which provides measurements of range-to-target and two components of angle-to-target with respect to the missile's body axes. A scenario which provided the most interesting computer simulation results was one that is sometimes referred to as a "difficult" engagement. The initial missile-target geometry is characterized by a small separation distance and a velocity orientation that requires the missile to operate at its maximum limits with very little margin for error. The initial conditions for this particular engagement were: a range-to-target of 4000 ft, an off-boresight angle (angle between the missile's velocity vector and its line-of-sight to target) of 40 degrees, and an aspect angle (angle between the line-of-sight vector and the target's velocity vector) of 135 degrees. The initial missile and target speeds are equal at 970 ft/sec (Mach \approx 0.9) and are both at the same altitude. A computer generated illustration of the encounter is shown in Fig. 5. The target's trajectory, in this figure, is depicted by the dotted line. During the first 2 seconds of the encounter, the missile is able to accurately estimate the position and velocity vector of the target, but because of the difficult initial geometry, can not achieve the acceleration to intercept the target. Its closest approach occurs at 2.5 seconds. Beyond this point, the target falls outside of the field of view of the missile's forward-looking radar. Ordinarily this would simply cause the missile to fly by the target. However, with the integrated strapdown design, the missile has accurately estimated its position and velocity vector relative to the target and continues to maneuver toward the target, intercepting it approximately 2.5 seconds later.

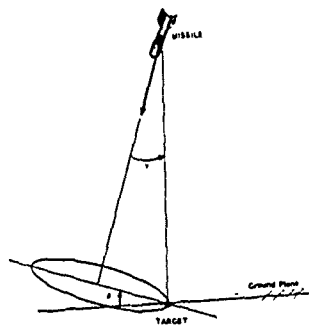
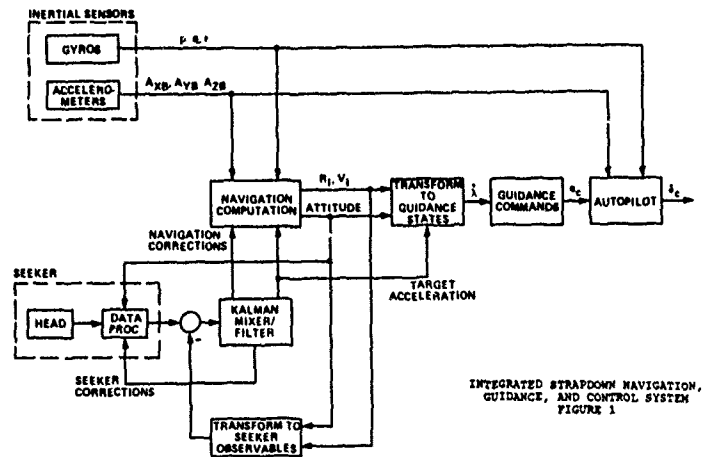
CONCLUDING REMARKS

A portion of the discussion on the benefits of using integrated strapdown avionics should include some aspects of implementation. Although all of our simulation studies have shown that comparable or superior performance to a conventional system could be achieved with a less expensive integrated strapdown configuration, in some cases it was necessary to augment originally programmed algorithms in order to compensate for certain unexpected effects that were discovered to be unique to strapdown systems. When one replaces more elaborate inertially stabilized instruments with their strapdown equivalents, the computer is required to perform the function of the hardware being removed (this, in fact, is one of the cost-saving features associated with the integrated strapdown approach). One must be careful, however, to make sure that the algorithm accounts for all factors (not always obvious), which under certain conditions can be sensitive to an instrument in a strapdown mode and yet insensitive to one stabilized inertially. For example, small strapdown seeker boresight errors (offset misalignments between the body-fixed coordinate frames of the strapdown seeker and the strapdown inertial platform) when coupled with high uncontrolled vehicle roll rates can seriously degrade performance. (On the other hand, seeker misalignment error in an inertially stabilized seeker does not interact with roll rate.) However, if one includes the boresight error as an additional state to be estimated by the Kalman filter, the problem is completely eliminated. Similarly, strapdown seeker and gyro scale-factor errors, if large enough and not included in the filter, can cause the closed-loop guidance and control system to become unstable. The level of scale-factor error causing the instability is a function of the characteristics of the closed-loop system (guidance law, autopilot, and airframe characteristics) [4]. While all of the weapon configurations in our investigation could adequately accommodate the instrument scale-factor errors, a more advanced guidance law/autopilot/vehicle configuration could present stability problems if the scale-factor errors are not properly compensated for in the algorithm.

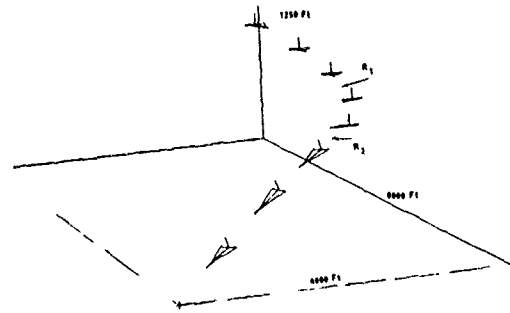
In summary, if all factors are accounted for, an extremely effective way of reducing the avionics cost of precision guided vehicles is to integrate the design of the various subsystems of navigation, guidance, and control with a common set of strapdown components and a Kalman mixer/filter. In addition to reducing the cost and improving the reliability of the overall system, proper mixing of the strapdown data can, in many cases, result in performance which is superior to that obtained with more expensive avionic components used in the conventional independent navigation, guidance, and control configuration.

REFERENCES

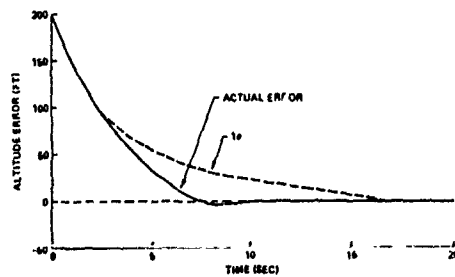
- [1] D.A. Haessig and B. Friedland, "Maximum Likelihood Estimation of Target Acceleration," *IEEE Conf on Decision and Control*, Las Vegas, NV, Dec. 12-15, 1984.
- [2] T.E. Bullock and S. Sangsuk-lam, "Maneuver Detection and Tracking with a Non-linear Target Model," *Proc. of 23rd Conf. on Decision and Control*, Las Vegas, NV, Dec. 1984.
- [3] D.E. Williams, J. Richman, and B. Friedland, "Design of an Integrated Strapdown Guidance and Control System for a Tactical Missile," *Proc. AIAA Guidance and Control Conf.*, Gatlinburg, TN, pp. 57-66, Aug. 15-17, 1983.
- [4] R.K. Mehra and R.D. Ehrich, "Strapdown Seeker Advanced Guidance," *Workshop on Bank-to-Turn Controlled Terminal Homing Missiles*, Laurel, MD, Sept. 19-20, 1984.



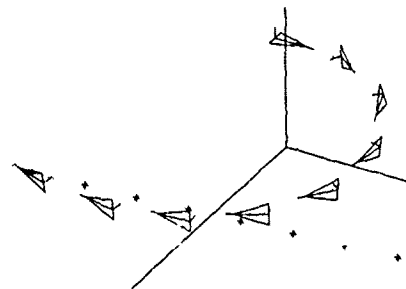
SENSOR GEOMETRY
FIGURE 2



TRAJECTORY OF VEHICLE TO TARGET
FIGURE 3



ERROR IN ESTIMATED ALTITUDE
FIGURE 4



TRAJECTORIES OF MISSILE AND TARGET
AIRCRAFT DURING AIR-TO-AIR ENCOUNTER
FIGURE 5

PART V

Integrated Navigation/Flight Control Systems

INTEGRATED NAVIGATION/FLIGHT CONTROL FOR FUTURE HIGH PERFORMANCE AIRCRAFT

by

Robert E. Ebner and A. David Klein
Litton Systems, Inc., Guidance and Control Systems Division
5500 Canoga Avenue, Woodland Hills, CA 91367-6698
United States

SUMMARY

Litton has delivered an Advanced Development Model (ADM) of an Integrated Inertial Sensor Assembly (IISA) on contract to the U.S. Naval Air Development Center. IISA is designed to provide all inertial sensor needs for modern military aircraft, including flight control and navigation, with reduced avionics cost through the use of redundant skewed inertial navigation sensors. Various design aspects of using six ring-laser gyros and six inertial-grade accelerometers in two, separated clusters are described. The redundancy management mechanization and the system design features for maximum flight safety are given. Navigation performance limits of strapdown INS, including the effects of skewed sensors, are presented. Laboratory testing will be performed by the Navy and flight testing will be conducted on an F-15 as part of a joint Navy/Air Force program.

INTRODUCTION

With the advent of the Ring Laser Gyro (RLG) and high-speed digital processing in a reasonable form factor, high-performance strapdown inertial navigation systems became feasible for use on high-performance military aircraft. The first generation of these systems is just now finding its way into military aircraft under the Navy CAINS II and USAF Standard Navigator programs providing 0.5 to 1.0 nmi/hr inertial performance. The full potential of these two technologies, however, lies in their application to provide multifunctional avionics capabilities. This has been recognized in the integrated systems architectures developed for both Navy advanced aircraft and the USAF Advanced Tactical fighter. Due to the extremely wide bandwidth inertial sensors, it becomes feasible to apply the same sensors used for inertial navigation to support the requirements of flight control and mission sensor stabilization. Such multifunction use of the system will reduce to a minimum the number of gyros, accelerometers and electronic components required to support these functions.

The accomplishment of this objective involves a great deal more than employing three or four conventional RLG inertial systems to provide the fault tolerance required. It requires attention to a set of parameters involving redundancy, noise and data latency.

IISA, which has been under development since 1982 sponsored by the Naval Air Development Center, is designed to satisfy the requirements for navigation, flight control and sensor stabilization with a minimum amount of hardware. It replaces other systems containing some 15 to 20 different gyros and accelerometers and provides fault-tolerant navigation and a fail-operational, fail-operational, fail-safe capability to support the requirements of flight control through its unique architectural approach. System hardware has now been delivered and will go through rigorous military flight testing under the ADA Based Integrated Control System III (ABICS III) Program being undertaken by McDonnell Aircraft Corporation under the sponsorship of the U.S. Air Force Flight Dynamics Laboratory. Flight testing will be performed on an F-15 aircraft configured with a dual redundant fly-by-wire system. This paper will review IISA characteristics and design features to support multifunction requirements and present the results of IISA laboratory and integration testing. The results of this program are critical to demonstrating the feasibility of this system concept for high performance aircraft.

SYSTEM DESCRIPTION

The inertial sensors of IISA are contained in two Inertial Navigation Assemblies (INA), each of which provides full, independent inertial navigation outputs. Two INAs have been delivered as part of the IISA program and two additional units were delivered for the ABICS III Program.

Within an INA, sensor axes are orthogonal but skewed relative to the aircraft yaw axis (see Figure 1). One accelerometer and one gyro in an INA are oriented along each skewed axis. Figure 1 depicts the orientation of axes when the INA are installed into the equipment bays of the aircraft. When one INA is installed into the right equipment bay, with 180° rotation about yaw relative to the identical left INA, the six sensor axes are then distributed uniformly about a 54.7° half-angle cone. No two axes are coincident, nor are three in the same plane. Thus, any three sensors may be used to derive three-axis outputs in aircraft axes after suitable computer transformation.

The two IISA units may also be installed into the same side of the vehicle, with the second unit rotated 180° relative to the first about a horizontal axis. The unit design must include provision for both nominal and upside-down installation, not included with the ADM design.

An INA is divided into three, largely independent channels. Each channel contains data from one gyro and one accelerometer plus related electronics, a preprocessor, provisions for output of data to the FCS and to the navigation computer, and independent low/high voltage power supplies. The navigation processor and its MIL-STD-1553B I/O are on the same power supply as one of the three sensor pair channels.

The packaging arrangement for the INA is shown in Figure 2. The three channels of electronics are physically separated to eliminate common failure modes. Wiring from the sensors to the sensor electronics is also kept physically separated to avoid short-circuit, EMI, etc., failure modes common to two channels.

Weight of the ADM inertial navigation assembly was 56 pounds. Based on use of more modern electronics, and conservative weight reduction techniques, the weight of a production model can be reduced to under 50 pounds. Power is approximately 100 watts.

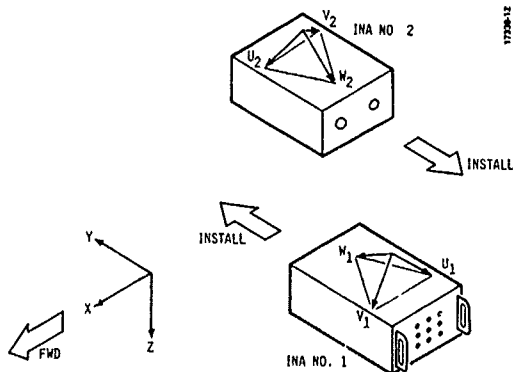


Figure 1. INA Installation Configuration

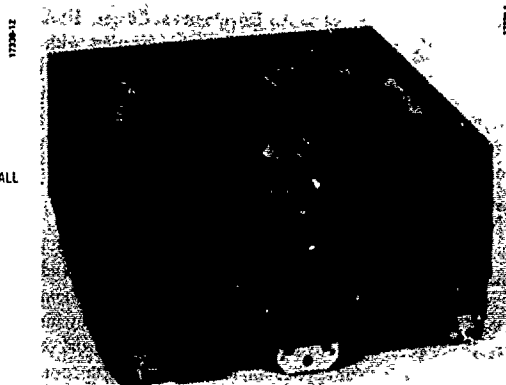


Figure 2. Inertial Navigation Assembly

NAVIGATION PERFORMANCE

The navigation performance requirements for IISA are similar to general, medium accuracy systems currently in inventory. Performance of strapdown inertial sensors using ring laser gyros has been described in the literature. A serpentine flight path, however, is not generally described. When the flight path varies back and forth repetitively through some significant angle using coordinated turns, navigation errors of a strapdown INS become strongly dependent upon gyro scale factor and axis alignment errors. This flight path is probably a realistic one for tactical aircraft during terrain avoidance and evasive maneuvering. Ring laser gyros can maintain excellent scale factor stability. Achieving the 1-2 arcsec axis alignment stability needed if a significant portion of flights is to contain serpentine maneuvers requires very careful design. On IISA, material selection and structural rigidity between gyros has been determined primarily to meet this difficult requirement.

Skewing of accelerometer axes requires that accelerometer scale factor stability be significantly better than for a nonskewed configuration. An accuracy requirement of 35 ppm scale factor tracking between the three accelerometers is within the capability of the accelerometers used on the ADM, as evidenced by long-term stability testing over a three year period.

Performance during vibration is essentially the same for skewed and unskewed sensors. As described in (1), gyro input axis bending is the major error source for strapdown navigators in a vibration environment. IISA has been designed for the most rigid gyro-to-gyro structure obtainable to attain accuracy goals during vibration.

The ADM IISA was flown in a company-owned Citation, and achieved a CEP position accuracy of 0.25 nmi/hr over five flights. Velocity accuracy was 1.54 ft/sec rms, per axis, at the end of 2-4 hour flights. A flight with 10 S-turns indicated that gyro scale factor error was 2.5 ppm and input axis alignment error was 0.42 arcsec. This clearly demonstrated the feasibility of precision inertial navigation performance using skewed sensors.

FLIGHT CONTROL SENSING

Performance

Inertial navigation gyros and accelerometers are orders of magnitude more accurate than those commonly used for flight control. Part of the accuracy is achieved by software modeling of residual errors and much of this benefit also applies to angular rate and acceleration outputs for flight control.

Software axis alignment correction, however, is more complex for a redundant system since it involves mixing of data between sensors. Also misalignments due to physical separation and vibration isolators cannot easily be compensated. Flight control accuracy requirements are limited, however. Therefore, full inertial-grade axis alignment accuracy is not provided for flight control sensor outputs.

The specified accuracy of outputs to the flight control system is shown below. Actual accuracy will be significantly better since the outputs are derived from inertial navigation grade sensors.

	Angular Rate	Acceleration
Scale Factor	0.1%	0.1%
Bias	1.5 deg/hr	4 mg
Alignment	1 milliradian	7 milliradians
Resolution	0.02 deg/sec	2 mg
Range	400 deg/sec	20 g

Time Delays and Synchronization

IISA is basically a digital sensor, to be used as part of a digital flight control system which is controlling the states of an aircraft in real time. Data sampling and processing time delays in the sensor element cause a destabilizing effect in an aircraft control system and must be carefully selected.

Gyro and accelerometer outputs consist essentially of pulse streams which are counted over some time interval to obtain an estimate of angular rate or acceleration. If this count interval is too long, extensive time delays are introduced into the FCS. Selection of count interval and subsequent digital filtering to reduce noise and quantization effects must be balanced against FCS delay and phase lag constraints.

Since the IISA sensor subsystem is implemented as six separate skewed gyro and accelerometer pairs, the data sampling intervals may begin at different times for each sensor, unless some form of cross-channel synchronization is employed. The primary effect of such a time-skew between sensors is to contaminate redundancy management sensor comparisons during very angular acceleration or rate of change of linear acceleration.

Data sampling is initially derived from a single clock in order to achieve required navigation accuracy. Each sensor pair separately monitors the accuracy of this clock, relative to its own. If an error is detected the sensor pair's clock is used. This leads to the asynchronous operation discussed above.

Anti-Aliasing Filters and Dither Noise

Modern flight control systems are digital and sensor data is sampled at some fixed frequency, e.g., 80 Hz for modern fighter aircraft. Sensor noise or vibration inputs at high frequencies can be aliased by the sampling process to a frequency within the flight-control bandwidth, causing control surface flutter or pilot discomfort. Therefore, it is common to filter gyro and accelerometer outputs to remove high-frequency noise. For digital sensors such as used in IISA, filters must be digital in nature and the sampling frequency must be greater than twice the highest noise or vibration frequency. Since IISA sensors are attached to vibration isolators, limiting sensed vibration bandwidth, digital filters iterated at 1 kHz will produce the required noise rejection.

It is desirable to reject noise in sensor outputs within 10 Hz of the FCS data sampling frequency and its harmonics. These are the frequencies which can potentially alias to the 0-10 Hz region, the maximum bandwidth of the FCS. This can be achieved, for example, by a low-pass filter. There is a tradeoff between filter noise rejection capability and time delays and lags which could potentially destabilize FCS loops.

The primary cause of acceleration noise is accelerometer quantization. Acceleration measurements are converted to digital form in the feedback path of the force rebalance loop. The quantization level of 0.0015 ft/sec at each end of the sample combined with the effective sample time of 0.017 sec produces a maximum quantization error of 0.18 ft/sec², leading very close to the measured 1-sigma value of 0.05 ft/sec². The gyro dither serves to randomize the error which would otherwise be a fixed-frequency for a given g-level. If an application requires a wider bandwidth with similar static acceleration noise, accelerometer quantization step size can easily be reduced. Gyro dither also produces acceleration noise due to the fact that accelerometers cannot be located at exactly the center of rotation of the instrument cluster. Nonlinearities are sufficiently small, however, so that no beat frequencies have been observed on acceleration outputs due to the fact that the three gyro dither frequencies are not identical.

Gyro noise is also primarily due to quantization and dither. A notch filter at gyro dither frequency effectively eliminates dither effects, and quantization of 0.46 arcsec reduces angular rate noise to under 0.03 deg/sec. If some anti-aliasing filtering is included, this can be reduced to well under 0.01 deg/sec.

Vibration Isolators

Inertial navigation systems, strapdown or gimballed, commonly protect the inertial sensors from the environment by using elastomeric vibration isolators. This is especially significant for shocks produced during handling by maintenance personnel, which are uncontrolled conditions. Another reason for vibration isolators is to protect the inertial sensors from military environmental testing. This is often orders of magnitude worse than the real aircraft environment, especially during the accelerated life test (endurance testing). Navigation performance in this artificial environment would also degrade without isolators. In a real aircraft environment, however, the aircraft equipment shelf provides significant reductions of high-frequency energy not accounted for in typical qualification test levels.

Flight control rate gyros and accelerometers are not commonly isolated from vibration. Use of an IISA mechanization introduces new design constraints which should be carefully analyzed in each specific application.

Vibration isolators form a damped, resonant system, with a typical amplification factor of 3-4. Thus, both linear and angular vibrational inputs are shaped by the isolator transfer function, with a tendency toward a fairly narrow bandwidth output. In addition, linear vibration inputs are transformed into angular motions at the sensor by unbalances and resonant frequency mismatches in the isolator system. This amplification and bandwidth-narrowing of vibration could cause problems in the flight control system if not properly considered.

The resonant frequency of elastomeric isolators varies significantly with temperature (+37%, -13%) and vibration amplitude (12% greater when input vibration at resonance is halved.). The angular-to-angular resonant frequency is nearly double that of the linear resonance. Thus, digital notch filters at isolator resonances would not be effective under all environmental conditions. Another approach is to locate isolator resonant frequencies above the anti-aliasing filter cut-off. This, however, leads to rather stiff isolators which could possibly become coincident with shelf-resonances. Since coinciding resonances are generally to be avoided (many inertial navigation error mechanisms, for example, vary with the square of vibration level), this approach is not recommended.

The IISA vibration isolator linear resonant frequency is being placed into the high-frequency band region of the flight control system, above the maximum frequency response of the FCS. Attenuation is provided by the FCS bending mode and anti-aliasing filters.

Redundancy Management with Skewed Sensors

Since the two groups of inertial sensors are on separate vibration isolation systems and are physically separated, accurate navigation cannot be achieved after a second failure of the same sensor type (one failure per group). Therefore, redundancy management is directed exclusively toward flight control requirements. Small sensor errors, which would degrade only strapdown navigation, are not detected internal to IISA.

The sequence of operations performed in IISA is illustrated in Figure 3. Sensor data is first reviewed for hard failures, detectable by normal self-test methods. The sensors themselves give an indication of failures through loop closure tests, loss-of-signal indications, etc. I/O tests assure that data has been correctly transmitted, and dynamic reasonableness tests detect spurious outputs inconsistent with the vehicle capability.

Due to the physical separation of the two sets of accelerometers, angular rotations and angular accelerations of the vehicle cause different accelerations to be sensed by each set. To allow direct comparison between acceleration measurements under dynamic conditions, each sensor output is related to a common point of the aircraft using the current best estimate of vehicle angular rate and angular acceleration along with known lever arm displacement from that point.

The six skewed gyro axes and six skewed accelerometers are spaced evenly on a 109.5° cone whose axis is vertical. Since no two axes are coincident and no three are in the same plane, full three-axis outputs can be provided with three failures of a sensor type. Reasonable geometry is available for any combination of failures, i.e., geometrical amplification of errors is less than a factor of 3.

Detection of up to three failures is assured by comparison of redundant sensor data in which are termed parity equations. These equations cancel vehicle angular rate, or acceleration in the case of accelerometers, and expose sensor errors. Because of information limitations, a third sensor failure of the same type can only be detected. Isolation of which of the four sensors active at that point has failed cannot be achieved except for hard failures which are detected by conventional self-test methods. For this reason IISA is termed fail-operational/fail-operational/fail-safe.

Six-gyro (similarly for accelerometers) parity equations can be formed by comparing each gyro output to a least-squares estimate of its output derived from the remaining sensors. Since there are always two sensors orthogonal to each axis, this results in six equations which are linear combinations of four sensor outputs. The orthogonal sensors cannot contribute to error detection. After sensor failures, a different set of parity equations is required. Again, linear equations involving four sensors can be formed, five equations after the first failure and only one after the second.

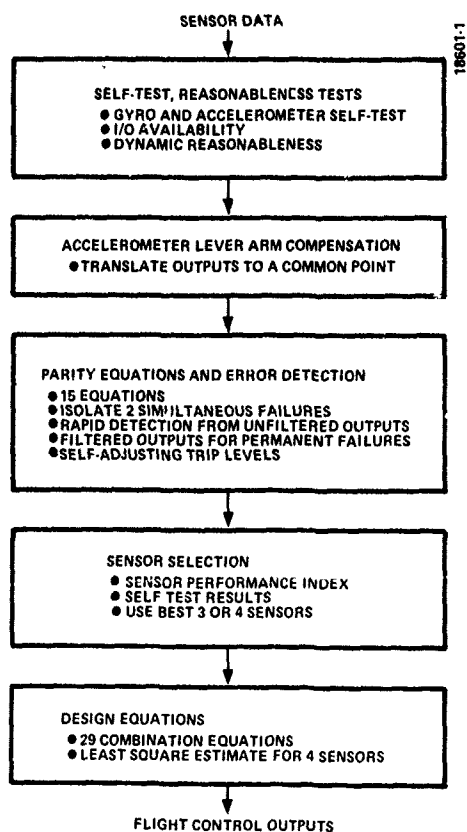


Figure 3. Redundancy Management Operation

With inertial navigation quality sensors, there is little value in combining data from all six sensors in a least-squares solution to derive outputs, rather than selecting a triad from a single unit, when available. Combining sensors simply adds another source of noise namely, the rocking motion of the second unit within the isolators. Therefore, whenever available, IISA outputs are derived from the three sensors on one unit. When there is one failed sensor in each unit, all four remaining sensors are used. For the condition where three failures are detected and failed sensors are known, the remaining three sensors are used. For the rare ambiguous case where all parity equations are failed and self-test cannot isolate the failure, warnings are issued to the pilot.

The equations which use selected sensor data to derive standard, orthogonal outputs to the flight control system are termed design equations. There are 29 sets of equations stored in the IISA computer, 20 for all the combinations of three sensors-at-a-time, and 9 for the least-squares estimates for four sensors-at-a-time, one failure in each unit. Only one set of design equations is used at a time, however.

Redundancy Management Performance

The quality of the redundancy management process rests on:

1. Noise level of parity equations
2. Thresholds that are used to detect failures
3. Transients that may occur in outputs when failures occur or if different sensors are selected due to normal noise conditions.

Realistic simulations have been performed to evaluate the effects of factors such as vibration isolators, anti-aliasing filters, and misalignments on the redundancy management process.

A parametric simulation has been performed based on a trajectory of a rapid roll into a high-g coordinated turn. The effects of isolator imbalance and mismatch, unit misalignment, and large installation lever arms were independently evaluated. Samples of the final computer run with all mechanisms present simultaneously are now presented. Table I shows the assumptions used in this simulation.

While the occurrence of two simultaneous failures appears extremely improbable from the standpoint of component reliability, sensors within a unit are under similar stresses (for example, local heating or shocks due to battle damage). Solution of all 15 potential parity equations during zero failure conditions, each derived from four sensors, allows detection and isolation of most soft dual-failure conditions, and this is the approach taken on IISA.

Under ideal conditions, parity equation outputs should be zero under any aircraft dynamic or vibration condition. However, because sensors are in separate, isolated units, shelf motion, isolator rocking and unit-to-unit misalignments cause parity equation outputs to appear when no sensor failures are present. For this reason, both for failure detection, the former: for detection of small, soft failures in some short time interval and the latter for very rapid detection of larger soft failures. The parity equation output level which trips gyro error detection logic is also varied as a function of angular rates and angular acceleration to avoid false alarms during maneuvers. A similar approach is used for acceleration trip levels.

The 15 parity equation outputs are scaled to be equal in their response to white noise from sensors. In general, however, all equations involving a sensor may not fail simultaneously. The parity equation coefficient for a given sensor, which is derived from the geometry, varies from equation to equation.

Thus, for a slowly degrading sensor, the 10 equations fail gradually rather than all at once. A sensor performance index (SPI) is formed for each sensor, equal to the number of parity equations it involves which have failed (0-10). The three sensors with the smallest SPI may be used for derivation of outputs. This is valid since in general three good sensors can be found easier than one bad one.

TABLE I
REDUNDANCY MANAGEMENT SIMULATION

- Aircraft velocity, 1000 ft/sec
- Rapid roll (800°/sec²) into 7-g coordinated turn
- Actuator response time, 0.05 second
- Lever arms
 - INA-1; X = -10, Y = -2, X = -1 ft
 - INA-2; X = 0, Y = 1.5, Z = 0.5 ft
- Vibration isolators
 - INA-1; 35 Hz resonance, 0.05-inch cg displacement
- Misalignments
 - INA-1; 0.2° yaw
 - INA-2; 0.2° roll
- Filtering
 - Gyro and acceleration measurements: 17 millisecond anti-aliasing filters
 - Parity equations: 25 millisecond low-pass filters

Parity Equation Outputs. The outputs from two gyro parity equations are shown in Figure 4 and are typical of the remainder. Equation Tij8 combines outputs from the U and W axes of each unit (Figure 1), while Equation Tij9 uses both U and V axes. The dominant error of Tij8 is the misalignment angle between units. The error of equation Tij9 occurs mainly from the vibration isolator frequency mismatch.

Accelerometer parity equation outputs (Figure 5) uses axes W₁ with U₂, V₂, and W₂ for Tij3, and V₁ with U₂, V₂, and W₂ for Tij2. Errors are mainly due to lever arm compensations and anti-aliasing filter lags.

Output Transients During Sensor Failures. During a soft soft sensor or sensor I/O failure, an output transient can occur due to: 1) use of different design equations involving a different set of sensors causing a change in the propagation of normal errors to the output, and 2) sudden removal of an acceptable soft error that may have been present for some time prior to finally exceeding the parity equation thresholds.

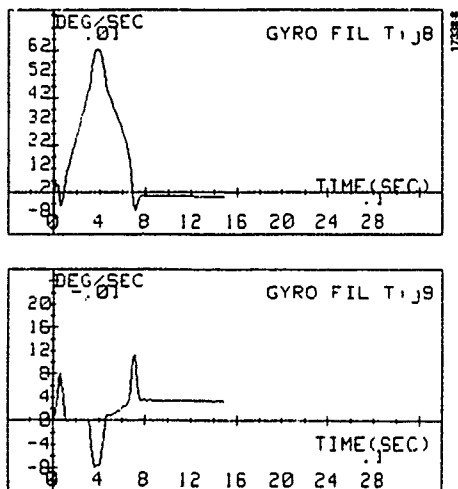


Figure 4. Sample Gyro Parity Equation Outputs

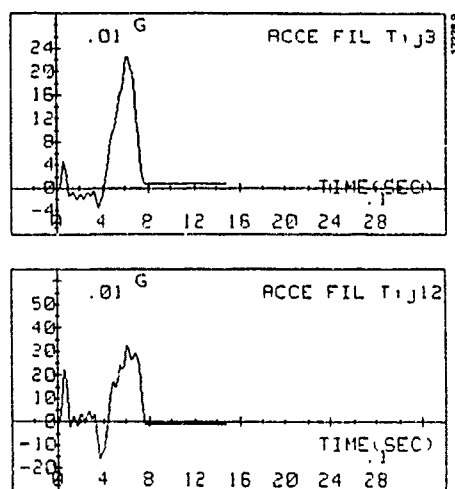


Figure 5. Sample Accelerometer Parity Equation Outputs

The former error was evaluated during the simulation by solving two sets of design equations simultaneously and differencing their results. The design equations selected are: one using two sensors in each unit, and the other using two sensors in the left unit and one sensor in the right unit. The differences between the two output computations are shown in Figure 6 for angular rate and Figure 7 for acceleration. These curves represent envelopes of possible transients where the actual transient depends on when, in time, the sensor failure occurred.

The latter error is a function of the failure thresholds used on parity equations. These thresholds will be made variable as a function of angular rate, acceleration, and lever arms. They will be set high enough so that false alarm error detections will be extremely improbable. Thus, transients following slowly increasing soft errors could be two to three times those indicated by Figures 6 and 7.

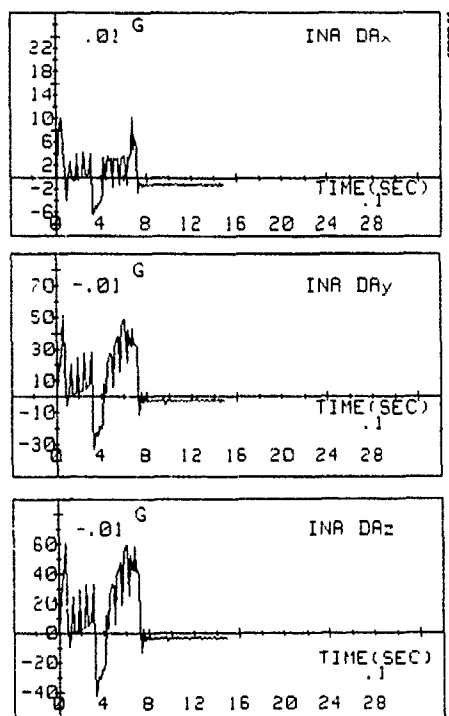


Figure 6. Maximum Acceleration Output Transient

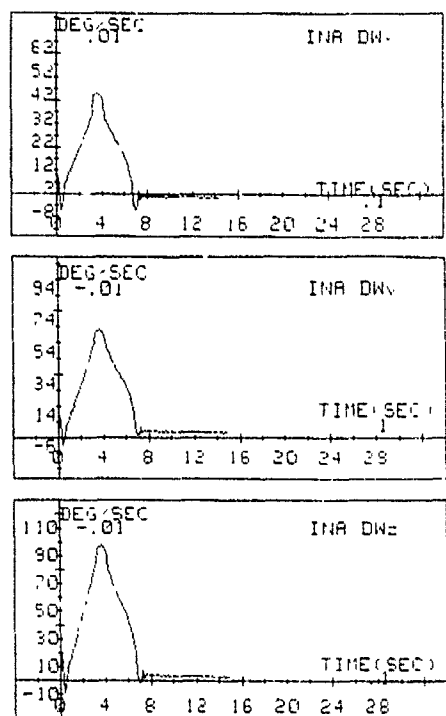


Figure 7. Maximum Angular Rate Output Transient

Simulation Conclusions. For the extreme maneuver that was simulated, the following maximum conditions occurred:

	Parity Noise	Switching Transient
Angular rate (degrees/sec)	0.6	1.0
Acceleration (g)	0.35	0.6

These levels are relatively small considering the high g-forces and transients from the maneuver itself, and should not lead to loss of control or pilot complaints.

The major source of these effects are misalignments and vibration isolators for angular rates, and lever arm compensations with anti-aliasing filter lags for acceleration measurements.

Perturbations of parity equations also occur due to vibration and fuselage bending modes. Simulations have not been performed at this time; however, the fuselage bending effects are expected to be quite significant for applications with widely separated sensor units. Methods such as described in Reference (2) may be needed for flight control compensation and to minimize parity equation noise.

Design Features for Flight Safety

The purpose of redundant system configurations is to achieve extremely low probability of functional failure. In the case of IISA, a failure of the sensor function can lead to loss of the aircraft since modern high-performance airframes are basically unstable, depending upon the FCS for stability. Component or software failure modes which cause simultaneous failure of two or more redundant elements severely degrade functional reliability and must thus be avoided.

Software reliability of an IISA is obtained by computational simplicity combined with extensive testing of the various software paths. The number of software paths for a sensor system is significantly smaller than for a modern flight control system, greatly simplifying the testing process. Therefore, some of the measures proposed for redundant FCS (such as different coding in each redundant computer) are not needed for an IISA. Extensive testing and rigorous control of changes are still needed, however, to eliminate the likelihood of a common failure of all computers during the same, probably unusual, condition.

Possibility of a common electrical failure mode is also eliminated by thorough design and testing, plus allowing only a bare minimum of communication between redundant elements. In the case of the Inertial Navigation Assembly, only a total of five complementary lines are allowed to cross module boundaries, three for serial data transfer of sensor data to the navigation processor and two for navigation data synchronization. In each case, failure propagation from one module to the next is prevented by isolating resistors and limiting diodes. In the case of the transfer of data from the sensors to the Digital Computer Assembly for redundancy management, isolation of serial data lines with separate drivers, resistors and diodes is again provided. Even under conditions of battle damage, a short to 115 V, 400 Hz will not propagate through circuitry to cause failure of more than one channel.

Dedicated serial data I/O is provided to transfer data from each of the sensors to the Digital Computer Assembly for redundancy management. A MIL-STD-1553 word format is used but since the transfer is not multiplexed between channels, full MIL-STD-1553 protocol is not required.

Redundant channels within a single unit are physically separated, so that under some remote failure condition, such as excessive EMI or heat-caused ejection of contaminants, failure mechanisms are contained within the channel and do not spread to other redundant elements. Dual cooling air ports are provided for each unit and each channel is driven by a separate power supply. Power supply wiring to the aircraft is also independent by channel for aircraft installation design flexibility.

TEST AND EVALUATION PLANS

The IISA ADM has been delivered to the U.S. Naval Air Development Center. Two additional INAs have been delivered to McDonnell-Douglas for flight test.

The flight tests will be performed on an F-15 aircraft at Edwards Air Force Base. Under a joint Navy/Air Force program, the two INA's and the CDU will be installed in the F-15. The INA's will be mounted with an approximate separation of three feet. The CDU will be mounted in the cockpit to provide system initialization and status and to introduce preplanned faults into the system.

IISA redundancy management software will be inserted within the F-15's Digital Electronics Flight Control System (DEFCS). The software will be rewritten by using the DOD High Order Language of Ada. Thus, IISA will closely resemble a production system.

The IISA test objectives and associated success criteria are five-fold: (1) IISA's airworthiness will be verified; the system must operate safely with no unacceptable system transients. (2) Aircraft flying qualities with IISA will be verified; IISA must not degrade aircraft stability and control. (3) IISA's susceptibility to false alarms (i.e., indication that a sensor has failed when in fact, it has not) will be examined; IISA must not be susceptible to false alarms. (4) IISA's redundancy management operation will be verified; faulty sensors must be identified and removed from the system in a manner that is transparent to the pilot. (5) IISA's navigation quality will be verified; terminal position and velocity accuracies must be at least as good as a medium accuracy navigation system.

SUMMARY

IISA has been designed to meet the flight safety needs of flight control inertial sensors while simultaneously operating as a medium accuracy inertial navigator. The system has been delivered and is currently undergoing test. Given the trends of the design of high-performance aircraft, an integrated system design such as IISA is essential for minimum avionics cost.

REFERENCES

1. J.G. Mark, R.E. Ebner, and A.K. Brown, "Design of RLG systems for High Vibration," PLANS '82 Symposium pp. 379-385.
2. W.K. Toolan and K.H. Grobert, "Development and Simulation Testing of an Integrated Sensory Subsystem (ISS) for Advanced Aircraft (Phase III)," Proc. 5th Digital Avionics Systems Conference, Seattle, Washington, November 1983.

SURVIVABLE PENETRATION

by

Carlos A. Bedoya, Gary N. Maroon and William J. Murphy, ScD
McDonnell Aircraft Company, McDonnell Douglas Corporation
Box 516, St. Louis, MO 63166, United States

and

Charles W. Chapoton, Jr., Ph.D
Texas Instruments

ABSTRACT

Threat densities expected on a modern battlefield do not allow penetrating tactical aircraft the option of simply flying around individual threats. As the threat becomes even more sophisticated in the 1990s new aircraft avionics systems will need to be fielded that will enable survivable penetration of tactical aircraft in an even more lethal threat environment. Recent advances in onboard mission planning, navigation, and terrain following/terrain avoidance/threat avoidance (TF/TA/TA) technologies and the onboard availability of stored digital terrain data enable the mechanization of such a survivable penetration capability.

Onboard mission planning constructs a survivable penetration reference corridor which takes into account terrain data, the location of known threats, the expected densities of unknown and mobile threats, and mission goals. An advanced aided navigation capability, using information from the global positioning system, aiding sensors such as radar, and terrain navigation features, is necessary to make maximum use of the onboard terrain data. The TF/TA/TA function computes flyable three-dimensional paths within the reference corridor accounting for aircraft performance limits, knowledge of the surrounding terrain, and information about the threat.

All three technologies use the Defense Mapping Agency's Digital Land Mass System (DLMS) terrain to provide look-ahead terrain masking and aided navigation. Through this survivable penetration methodology, advanced tactical aircraft can have enhanced aircraft survivability.

1. INTRODUCTION

An improved penetration capability is necessary to enhance the survivability of tactical aircraft in dense threat environments. Because of significant advances in Soviet air defense technology, sophisticated defenses will be deployed in such numbers that conventional penetration and attack profiles could result in unacceptable losses. In addition, tactical aircraft must retain their effectiveness at night and in inclement weather. Thus the need is paramount for an integrated system that provides survivable penetration through an onboard mission planner, a precision all-weather navigation capability, and a TF/TA/TA function.

The survivable penetration system's onboard mission planning function will determine the best corridor (most survivable within specified constraints) from one point to another. Typically the penetration corridor is defined by a set of waypoints. It starts at the forward edge of the battle area (FEBA), proceeds to the target area, and returns back from the target area across the FEBA. Factors influencing the choice of the penetration corridor include threat distribution, restricted areas, terrain, aircraft performance, available fuel, and time. Survivability is further improved either by very low altitude TF/TA/TA to take best advantage of local terrain masking to threats or by high altitude, high speed flight to stay outside surface-to-air missiles (SAM) and antiaircraft artillery (AAA) envelopes.

Accurate navigation is a fundamental capability needed for the survivable penetration system. Accurate knowledge of the tactical aircraft's location and velocity is key to onboard mission planning and TF/TA/TA. Accurate position information allows onboard mission planning to account for a known threat's masking envelope. During TF/TA/TA flight, accurate knowledge of the aircraft location relative to onboard digital terrain elevation data (DTED) enables terrain following flight over the DTED with intermittent operation of a forward looking terrain following radar. This accurate navigation capability can be performed through a moderate cost/performance inertial navigation system (INS) integrated with information from other available sensors. These aiding sensors would include, but not be limited to, doppler radar, a global positioning system (GPS), and terrain aided navigation that uses radar altimeter measurements and the onboard DTED.

TF/TA/TA will provide the penetrating tactical aircraft with a survivable three-dimensional flight path within the corridor computed by the onboard mission planner. The penetrating aircraft's improved probability of survival is achieved through stealth and tactical surprise. The aircraft can covertly penetrate the air defense threats by minimizing the onboard rf emissions from a terrain following radar. Survivability is enhanced by masking the penetrating tactical aircraft from the threats. Denying the threat direct line of sight by flying at very low altitudes minimizes the aircraft's exposure thus increasing aircraft survivability. Current day terrain following systems use a forward looking radar and algorithms to permit safe 50-meter above ground level (AGL) flight. With the use of an onboard DTED database, new approaches to local flight path planning are enabled. Today's data processing resources allow for flight planning algorithms that can compute three-dimensional flight paths.

The availability of DTED also allows the system to "see" through a hill and plan a trajectory beyond a forward looking radar's line of sight. Accurate navigation allows the radar to be turned completely off. The aircraft would compute its flight path directly from the DTED. In addition to this silent terrain following capability, the DTED enables lateral path generation that would include terrain avoidance (going around a hill) and threat avoidance (avoiding direct exposure to a threat).

2. SURVIVABLE PENETRATION REFERENCE PATH

Techniques have been developed to generate the survivable penetration reference path/corridor through use of an onboard mission planner, Reference 1. The survivable penetration corridor that is defined by the reference path and the corridor width is input to the TF/TA/TA function. TF/TA/TA further refines the path to be flown by accounting for detailed aircraft performance characteristics and threat information within the corridor.

This reference path is the one with lowest "danger" that incorporates knowledge about known threats and threats that are either mobile or whose location is totally unknown. For known threats, techniques were developed that either avoid these locations or use local terrain masking to minimize threat encounters. For threats which are unknown (or mobile) or for threats with overlapping coverage, it is important to use a mathematical optimization method (such as, dynamic programming) that effectively minimizes the threat exposure for the expected distribution of threats.

In developing a survivable penetration path trajectory it is necessary, in addition to the optimality criteria, that the chosen path be tolerant of threat and terrain data-base inaccuracies, accept pilot interactions, and be insensitive to minor path departures. Thus, the uncertainty of the data must be consistent with the character and precision of the trajectory. Small departures from nominal trajectories which achieve optimal performance by flying through the "eye-of-the-storm" may incur greater operational penalties than departures from a nominal trajectory which has slightly less optimum performance but which tolerates a wider corridor around the nominal. This "robust" characteristic is an important ingredient of penetration trajectories. Methods of incorporating this quality have been developed and depend on sensitivity studies of optimality criteria as well as imposed trajectory constraints.

2.1 SURVIVABLE PENETRATION METHODOLOGY

The optimization algorithm is a two-part process, as described in Figure 1. Dynamic programming generates a reference trajectory which defines the optimal corridor within which the aircraft is free to pick its own path. This reference trajectory is then refined using a TF/TA/TA algorithm. The output of the TF/TA/TA algorithm is a flyable trajectory which seeks a refined aircraft trajectory within the reference corridor.

Mission goals and constraints along with pilot inputs are used to define mission requirements. The reference trajectory generator makes use of DL'S (Digital Land Mass System) terrain data and current threat scenario information, based on a one to two nautical mile (nm) grid size. A TF/TA/TA altitude response model is used to predict the expected average aircraft altitude based on velocity, terrain standard deviation, and the pilot-selected terrain clearance setting. The reference trajectory is defined in steering buffers which can be input to the TF/TA/TA.

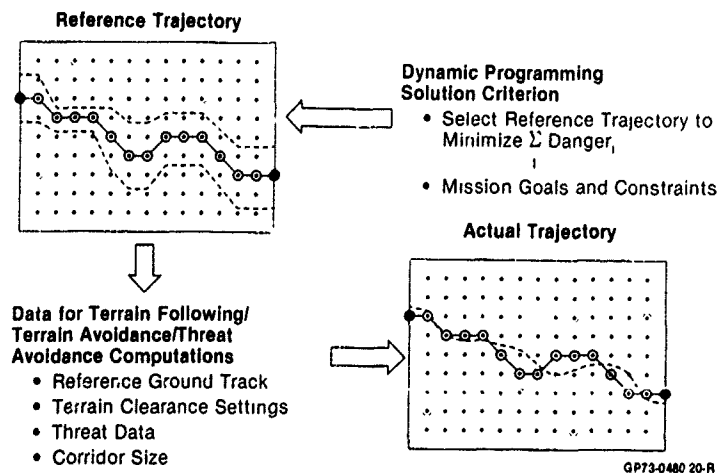


Figure 1. Survivable Penetration Methodology

The dynamic programming solution contains all the optimal paths from any point in the mission area to the target or goal. This allows a new reference trajectory to be immediately available should the pilot depart from the current reference trajectory. A new dynamic programming solution will be computed when the aircraft receives a mission redirect. A new target has been selected or the threat scenario is updated.

2.2 REFERENCE TRAJECTORY GENERATION

Dynamic programming is a recursive, discrete optimization process. Inputs to the optimization process include: mission description, optimization cost function, and optimization space. The mission description provides the dynamic programming process with a waypoint description of the mission in terms of goals such as attack initiation points, optimization boundaries, and other pilot-provided information. The cost function is a measure of potential threat encounters. A Lagrange multiplier is used to meet time and fuel constraints. The optimization space is a discrete state space which defines all allowable aircraft states and transitions between states. The use of a 1 to 2 nm grid size, along with the constraint of no more than a 45 degree change in heading allowed from one state to the next, provides a feasible corridor. One step of "previous heading" information is incorporated into the dynamic programming solution by including aircraft heading with lateral position in the definition of the state space.

Dynamic programming affords an efficient methodology to perform an exhaustive search over the entire optimization space. The recursive nature of dynamic programming allows the optimal path from each state in the space to the goal to be calculated in far less time than the optimal path from a specified start state to the goal could be calculated in a brute force manner.

The reference trajectory should be "robust" in the sense that database inaccuracies and pilot- or TF/TA/TA-imposed excursions about the path will not invalidate the solution. A solution tolerant of real-world interactions is preferred over a solution which requires the pilot or the system to continually "thread the needle" in terms of meeting position constraints. With a nontolerant solution the probability of straying off the reference path and onto a highly dangerous region is much greater.

2.3 REFERENCE TRAJECTORY PERFORMANCE MEASURE

The performance measure (optimization cost function) for the reference trajectory problem is:

$$\min \sum_i (D_i + Ct_i) \cdot \Delta t_i$$

where D_i is a measure of danger in terrain cell i , hereafter referred to as the "danger index", and Ct_i is a Lagrange multiplier representing the relative cost of time/fuel.

Sources of danger considered by the optimization procedure are known stationary threats, known mobile threats and unknown threats. Known stationary threats have been identified by intelligence data by type and position and are expected to remain stationary during the mission. Known mobile threats are represented by a zone or region within which the presence of mobile threats is highly probable. Unknown threats are represented by a region in which threats are considered to be randomly located with some nonzero probability. The unknown threat concept allows the system to take maximum advantage of the terrain in the absence of known threats. The danger of each individual threat is defined in terms of the threat's lethality (or effectiveness) and the expected terrain interaction. Updated threat information will cause the optimization software to generate a new solution.

The discrete nature of the performance measure and current threat modeling capabilities do not justify the complexity required to use probability of kill (or survival) estimates as part of the danger index. An absolute measure of danger is not required to find a good path. What is required is a relative measure of danger. Estimates of the actual survivability, if required, can be made using operational analysis survivability models. The system has been designed to use stationary threat effectiveness measures defined in terms of along-track and across-track distances. Mobile threat effectiveness is included by using a measure of the expected threat effectiveness based on the threat type and density. Unknown threat costs are included by assigning a constant cost to the entire flight radius of the aircraft beyond the FEBA (Forward Edge of Battle Area).

2.4 GENERATION OF THE DANGER INDEX

The danger index is a linear combination of the computed costs from the three threat classes (known, mobile, and unknown). The combination process is illustrated in Figure 2. The effects of terrain masking are taken into account using two methods: line-of-sight masking and area intervisibility calculations.

Line-of-sight calculations define the masking map for each known stationary threat. The effects of terrain on known mobile and unknown threats are taken into account using a nondirectional measure of masking effectiveness called "area intervisibility". Area intervisibility is a measure of the probability that a randomly located threat size has a clear line of sight to the aircraft.

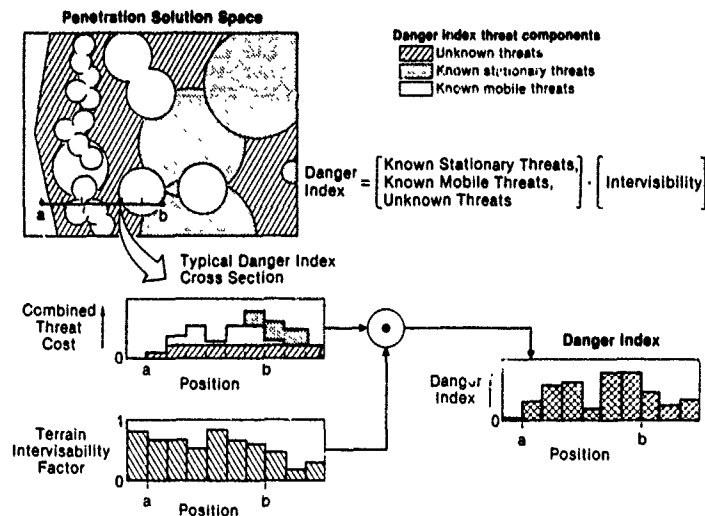


Figure 2. Danger Index Computation

Known threats that are stationary have effectiveness templates that are directional in nature. Provisions are made in the onboard mission planning algorithm to account for stationary threat lethality as a function of threat type and corresponding characteristics, aircraft type, the along-track and cross track aircraft position relative to site, aircraft speed and direction, as well as the difference in altitude between site and aircraft.

The template model used for the system checkout and demonstration defines effectiveness in terms of range from the site. The range is found using the along- and across-track inputs. Threat characteristics accounting for minimum and maximum range and altitude are also taken into account. A typical threat template for a known fixed threat is shown in Figure 3. A single effectiveness contour is depicted there and, as can be seen, is a directional function of down-range and cross-range distance coordinates and relative elevation of the target aircraft with respect to the threat. Since no threat position will be known exactly, some smearing of the threat template is desirable. A typical smeared template is also shown in Figure 3. Threat smearing is only applied to threats with location uncertainties typically less than 1 nm. For location uncertainties any greater than 1 nm, the threat is considered as a known mobile type and area intervisibility rather than geometric line-of-sight (terrain masking) would be used to account for terrain effects.

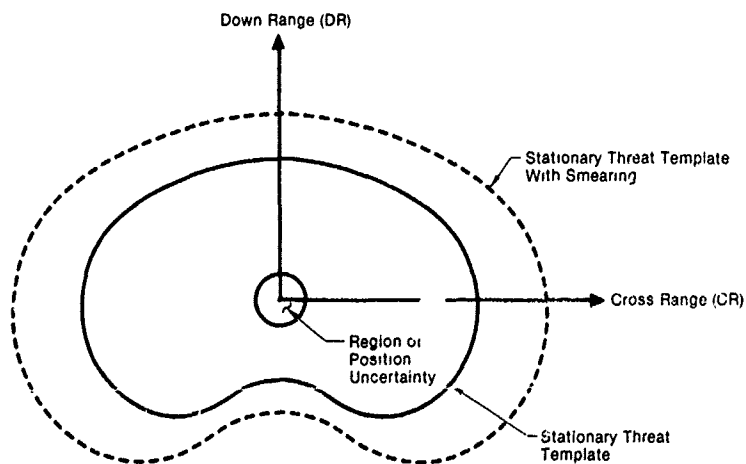


Figure 3. Typical Stationary Threat Template

The inclusion of mobile threats and unknown threat costs into the onboard mission planning process has created a need for a way to include the effects of terrain on loosely defined threats. The concept of "area intervisibility" has evolved to fill this need. Area intervisibility is the probability that a potential ground based threat site randomly located within a specified look radius (ground grange) from an aircraft at a given altitude AGL, has an un/interrupted line-of-sight to that aircraft. The definition of terms and the geometry of area intervisibility is shown in Figure 4.

Each Point on an Area Intervisibility Map Indicates the Probability That a Threat Site, Randomly Located Within a Given Radius of That Point, Will Have an Unmasked Line-of-sight to an Aircraft at a Specified Altitude (AGL) Above That Point

The Values in Any Single Intervisibility Map Are a Function of the Specified Aircraft Altitude and "Look" Radius

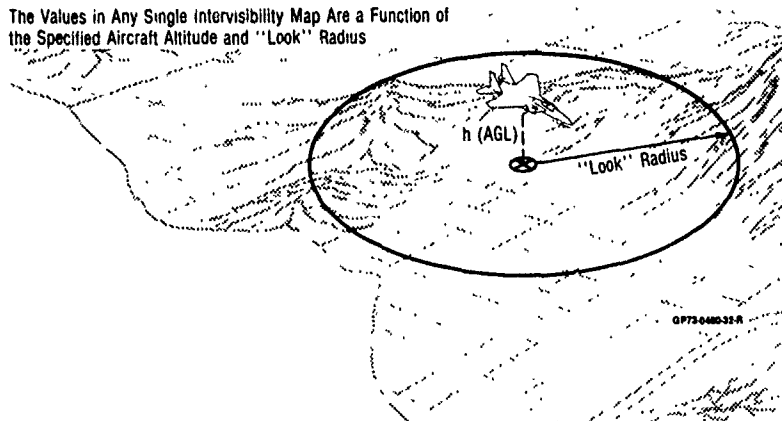


Figure 4. Area Intervisibility Definition

Figure 5 shows an example of a patch of DLMS Level 1 terrain data near Fulda, Germany. The altitude data has been quantized into eight bins, gray being the lowest and white the highest in altitude. Also shown is the area intervisibility for this terrain quantized from low (gray) to high (white). In comparing terrain altitude to intervisibility, an initial observation can be made that low altitudes are not necessarily areas of low area intervisibility. An obvious example would be a wide valley or a flat plain; in these instances a large number of neighboring points will have unmasked lines-of-sight. Often, areas of higher altitude but rougher terrain will offer lower values of area intervisibility.

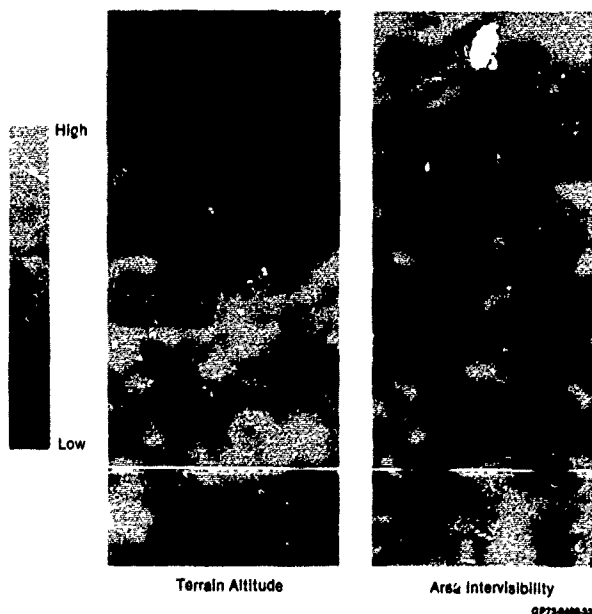


Figure 5. Terrain and Area Intervisibility

The processing requirements for the intervisibility calculations are high. Fortunately the processing can be done on the ground prior to the actual path optimization process. Area intervisibility as defined is only a function of the terrain and aircraft position and not of threat placement. The area intervisibility maps need only be calculated once for a set of different altitudes.

2.5 AREA INTERVISIBILITY ANALYSIS

The look radius used in the intervisibility calculations, as defined in Figure 4, affects the nature of the intervisibility maps. The expected value of the area intervisibility is driven by the chosen look radius. An extremely short look distance (on the order of 5-10 DLMS data points) will, in general, cause the calculated value of intervisibility to be very near one for all points, yielding very little information about the terrain. An extremely large value for the look radius will result in a value of intervisibility near zero for all cases because of the large number of points far away from the aircraft which are most apt to be masked.

Figure 6 demonstrates the effect of look radius on the intervisibility calculation. Area intervisibility calculations were made using second-order statistical terrain. The correlation distance of 10,000 ft and 500 ft standard deviation parameters used to generate the terrain are typical of the Fulda area of Germany. The right-hand column of data shows the portion of terrain used for the calculation and six intervisibility maps calculated using the indicated look radius. The calculations were made using an aircraft altitude of 200 ft. AGL. Area intervisibility is presented on an absolute scale from dark grey representing 0-12.5 percent intervisibility to white showing 87.5 to 100 percent intervisibility. Look radii at both the extremes exhibit a good deal of saturation at the maximum and minimum area intervisibility values respectively. Figure 7 shows a plot of the expected value of area intervisibility along with the minimum and maximum values encountered as a function of look radius. Look radii of 3 and 5 nm show a large unsaturated range of intervisibility values.

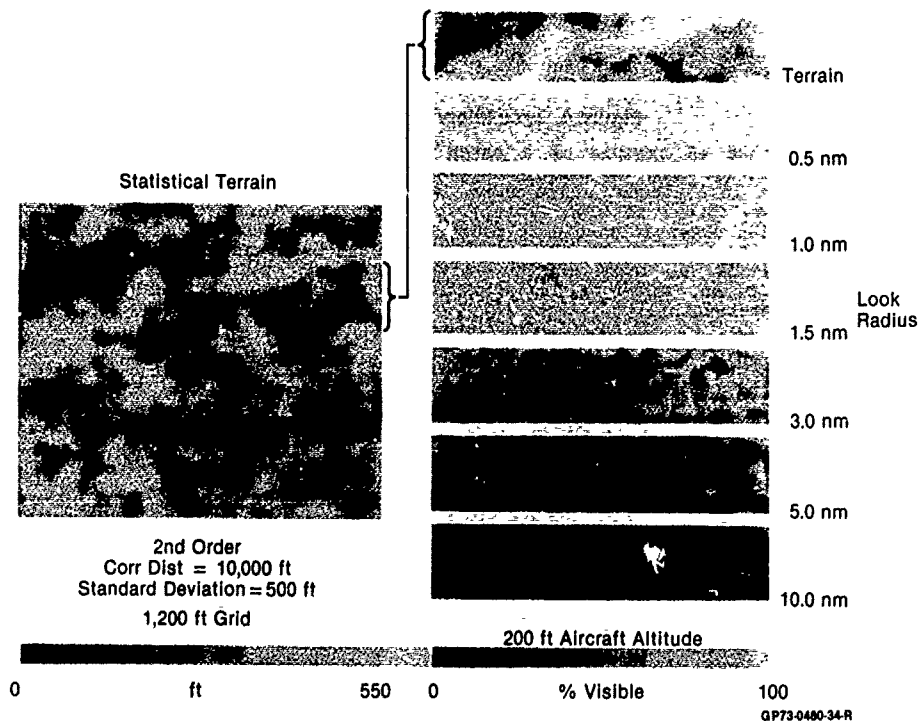


Figure 6. Area Intervisibility - Look Radius Sensitivity

The effects of altitude were evaluated by setting the look radius to 3 nm and calculating area intervisibility maps at altitudes ranging from 0 to 1000 ft. AGL. Figure 8 presents samples of altitude runs calculated using the same statistical terrain data. The calculated maps exhibit a rich range of values with very little saturation of data over the altitudes of interest for TF/TA excursions. A plot of the expected value of area intervisibility as a function of altitude is shown in Figure 9. The curve is very closely approximated by an exponential form. Expected area intervisibility values calculated using actual DLMS terrain data in central Germany show a high correlation with the statistical terrain calculation, as shown by the actual terrain values plotted on Figure 9.

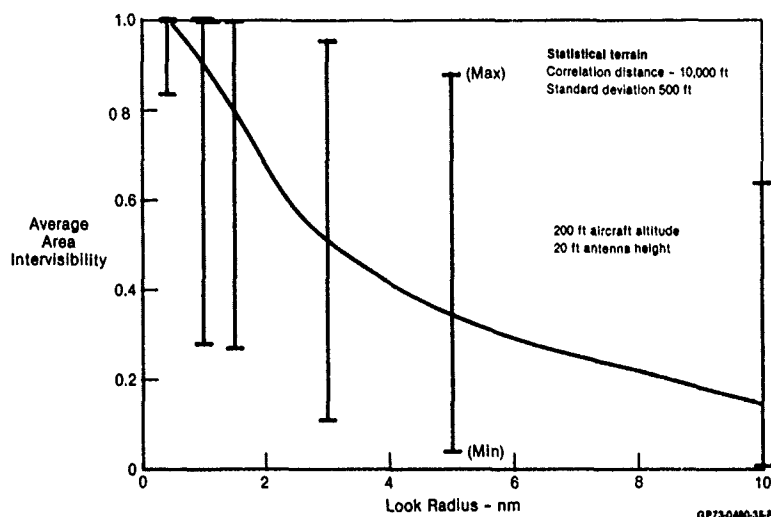


Figure 7. Average Area Intervisibility vs Look Radius

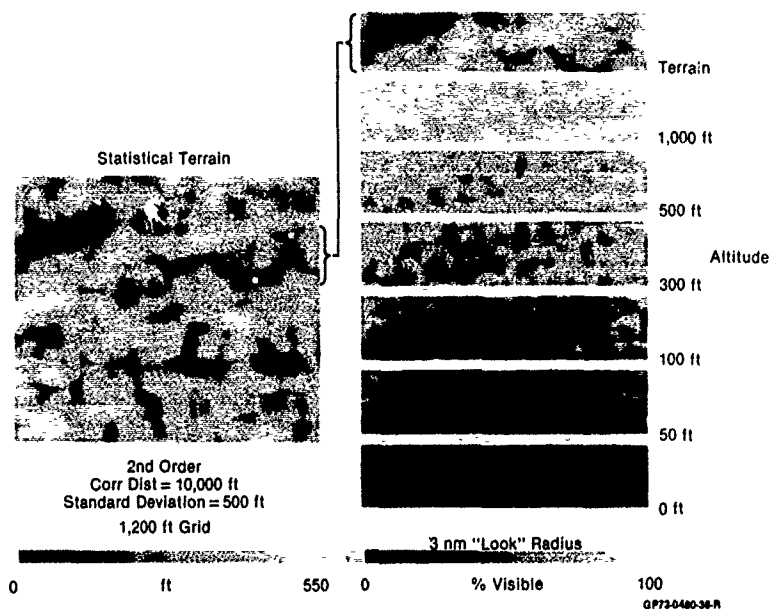


Figure 8. Area Intervisibility - Altitude Sensitivity

The value of look radius used for area intervisibility calculations can be varied as a function of aircraft altitude, or it can be held constant for all calculations. The look radius could be defined to approximate the aircraft horizon as a function of aircraft altitude. In this latter case, the area intervisibility calculation will include the set of all ground sites that could have line of sight to the aircraft. One problem that occurs in such a scheme is the tendency for the intervisibility measure to decrease as the aircraft altitude increases. Such a trend is clearly undesirable. The definition of intervisibility could be changed to account for the number of ground sites that have line of sight rather than the probabilistic definition. Area intervisibility in this case would be constrained to be monotonic increasing. For an aircraft traveling at 300 ft AGL, the "bald" Earth horizon is approximately 18.5 nm away. In analysis runs, the use of long look radii had an averaging effect on the intervisibility values thus decreasing the available information. In these cases, a driving factor in the intervisibility measure is the large number of points near the outside of the search area.

It was found that the use of a constant, a moderate value of look radius enhances the validity of the area intervisibility measure. For system demonstration a look radius of 3 nm was used. This value takes advantage of terrain information nearer to the aircraft, which has a greater probability of affecting the aircraft.

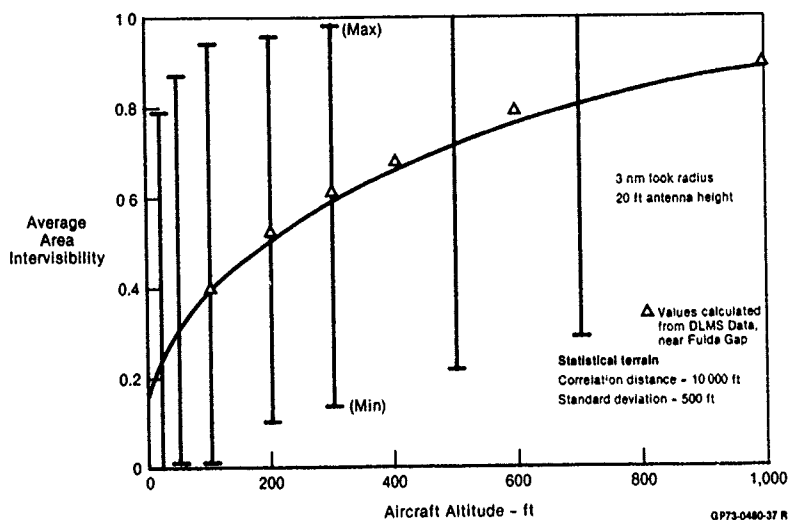


Figure 9. Average Area Intervisibility vs Aircraft Altitude

2.6 REFERENCE TRAJECTORY EXAMPLE

An example of a dynamic programming solution is shown in Figure 10. Optimum reference trajectories originate on the left side of the map and travel to a goal located at point A. The cost map was based on area intervisibility and was generated from DLMS data near the Fulda area of southern Germany. The map is approximately 38 nm x 30 nm with the lower left corner located at 50° 30' latitude and 9° 30' longitude. The grid size used was 1 nm square. The dynamic programming solution was restricted to a maximum turn of 45° at any one state. The dark shade represents low area intervisibility and the light represents areas of high intervisibility. The same reference trajectories are superimposed on the original DLMS terrain altitude data in Figure 11.

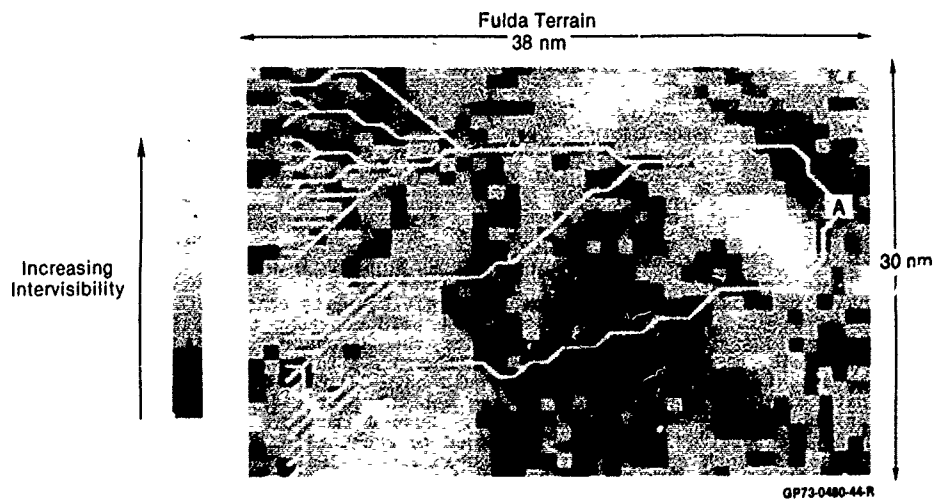


Figure 10. Optimal Paths
Intervisibility Coarse Grid

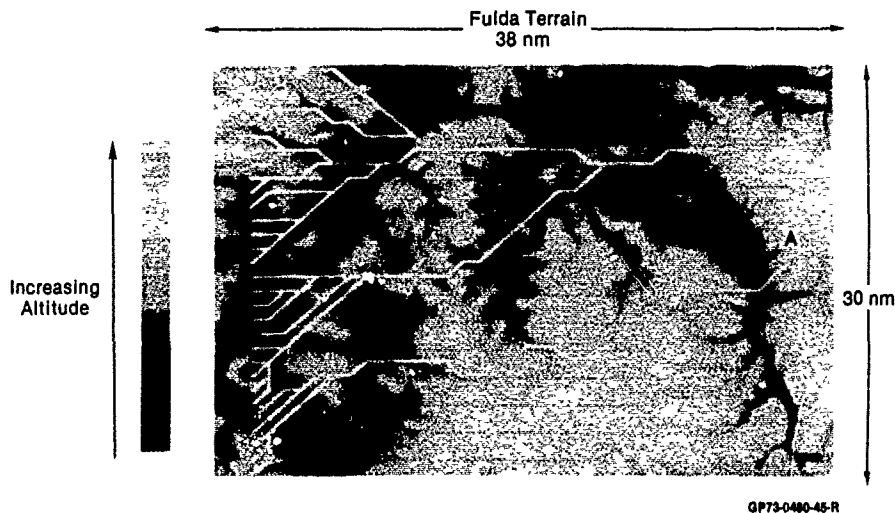


Figure 11. Optimal Paths
Terrain Coarse Grid

3. NAVIGATION

Survivable penetration missions require very accurate knowledge of the penetrating aircraft's position and velocity. Navigation accuracy is a key factor if successful low altitude missions are to be undertaken. The accurate navigation system aids the air crew in at least three ways. First, it provides the key ingredient to situational awareness, current location. This is especially important under conditions where tactical considerations have indicated a change in planned routing. Secondly, it reduces the crew workload. Interviews with tactical aircrews indicate that over 50% of their time in a low level interdiction is spent managing a navigation system that requires frequent manual updating, monitoring and keyboard insertions. Finally, there is the crucial issue of precision cueing in the target area that a good navigation system can give in terms of a lead-in to planned targets.

Aided Navigation - In the late 1950's inertial navigation systems (INS) using gyros and accelerometers for determining position and velocity were introduced into tactical aircraft. These systems provided an autonomous source of aircraft attitude, position and velocity that was used for aircraft guidance and weapon delivery. Although great improvements in the accuracy and reaction of these systems has been achieved over the past two decades (See Figure 12) the aircraft's mission, weapon delivery accuracy and sensor cueing requirements have begun to exceed the capability of current INS's of moderate cost and performance (~\$150K, 0.8 NM/HR).

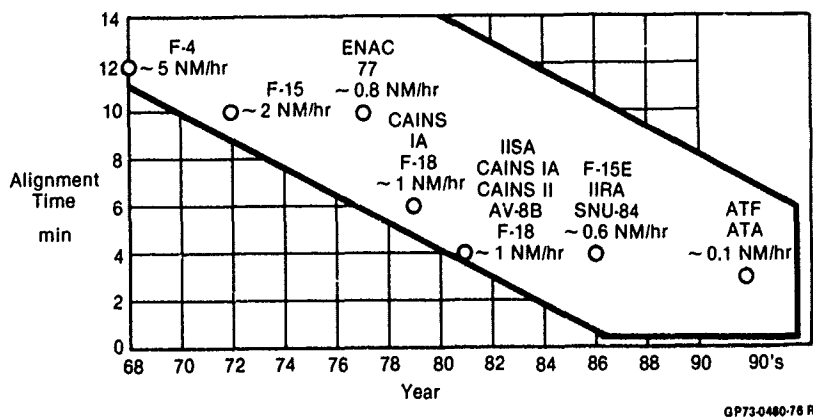


Figure 12. INS Performance Trend

There are many sources of navigation information in the sensor suites of modern tactical aircraft. Figure 13 lists typical sensors and the information available for use by the aircraft navigation system. Note that the primary function of a sensor need not be for navigation (i.e. the radar altimeter); yet it provides information that can be exploited in an aided navigation system.

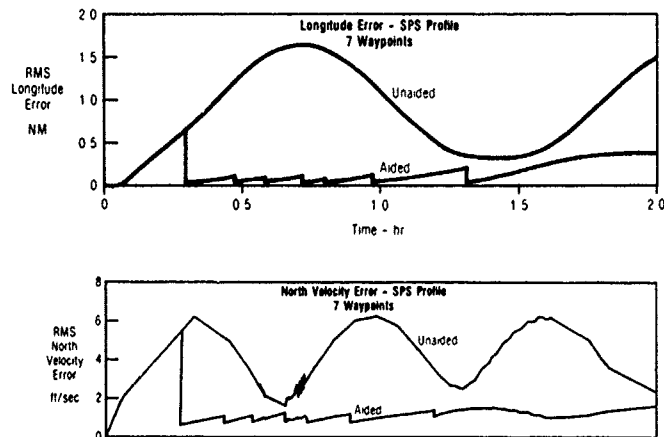
Sensor	Position	Velocity	Altitude	Heading	Time	Range	Attitude	Acceleration	Angular Rates
INS	✓	✓	✓ (1)	✓			✓	✓	✓
GPS	✓	✓	✓		✓				
JTIDS	✓ (6)		✓		✓				
Radar	✓ (5)	✓							
Barometric Altimeter			✓ (1)						
Radar Altimeter			✓ (6)						
Air Data Computer		✓ (2)	✓ (1)				✓ (3)		
Flight Control Sensors								✓	✓
AHRS				✓ (4)			✓		
FLIR	✓ (5)	✓					✓ (6)		
LANTIRN	✓ (5)	✓	✓ (6)				✓ (6)		
Laser Rangefinder/ Target Designator	✓ (5)	✓	✓ (6)				✓ (6)		
TACAN	✓			✓ (6)			✓ (6)		
Beacon	✓ (5)			✓ (6)			✓ (6)		
CO ₂ Laser	✓ (5)	✓	✓ (6)						
DLMS Data/ (Correlation Algorithm)	✓		✓						
MAD				✓ (4)					
Helmet Sight	✓ (5)								

Notes
 (1) Barometric
 (2) Airspeed true and indicated
 (3) Angle of attack
 (4) Magnetic
 (5) When an update location is known (fixed ground point)
 (6) Relative

GP73-077-R

Figure 13. Navigation Data

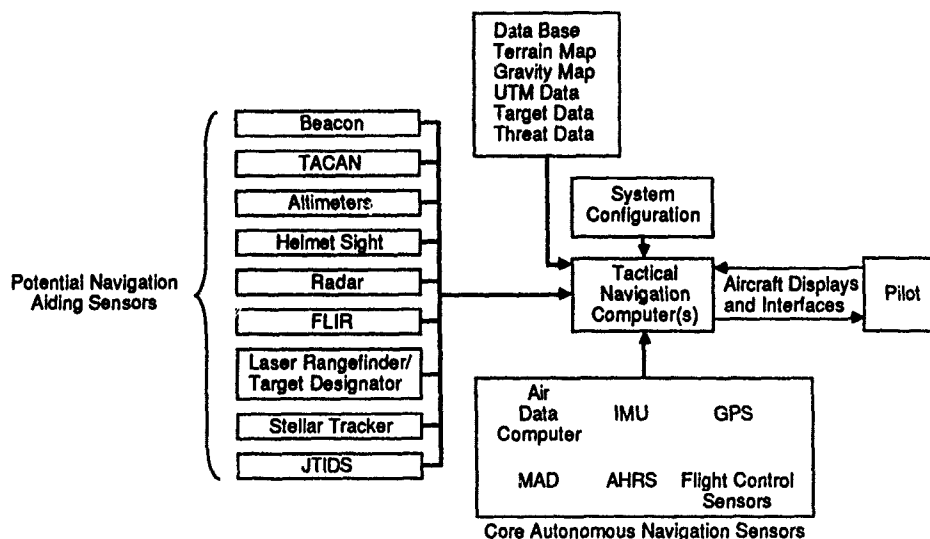
A typical aided navigation system consists of a medium accuracy INS (using information from other available sensors, such as a doppler radar), a synthetic aperture radar (SAR), visual target designations, a tactical air navigation set (TACAN), electro-optical sensors, etc. to update its position and velocity errors. Figure 14 illustrates the improvement which can be obtained with a medium accuracy INS when aided by a SAR in a typical aircraft mission.



GP73-0480-82 R

Figure 14. Synthetic Aperture Radar Aiding

Architectures - A candidate architecture of an aided system is presented in Figure 15. Various mechanizations are viable with the simplest approach being the periodic resetting of the INS to the value measured by the aiding sensor. The disadvantage of this approach is that the error sources in the INS and the updating sensor are not corrected and, therefore, the autonomous INS operation is not improved and the INS is not calibrated by the measurement.



GP73-0480-47-D

Figure 15. Aided Navigation Architecture

The use of Kalman filtering provides many desirable features. With proper mechanization and error modeling, INS alignment errors can be removed and source error calibrations can be performed on the INS instruments as well as in the aiding sensor. For example, if a doppler radar is used to provide accurate velocity information, INS gyro and accelerometer biases, scale factors and misalignments can be estimated as well as radar boresight and scale factor errors. If the INS were then required to perform autonomously, all bias and systematic errors would be compensated. Likewise, subsequent radar velocity measurements would be more accurate since its boresight error and scale factor errors would be removed.

Many Kalman filter architectures exist. The literature discusses the features and benefits of using a single central Kalman filter with raw sensor data as inputs, a distributed set of multiple Kalman filters, the use of a standard Kalman filter, an open or closed loop Kalman filter. Suffice it to say that many viable architectures are possible in implementing the aided navigation system using Kalman filtering techniques.

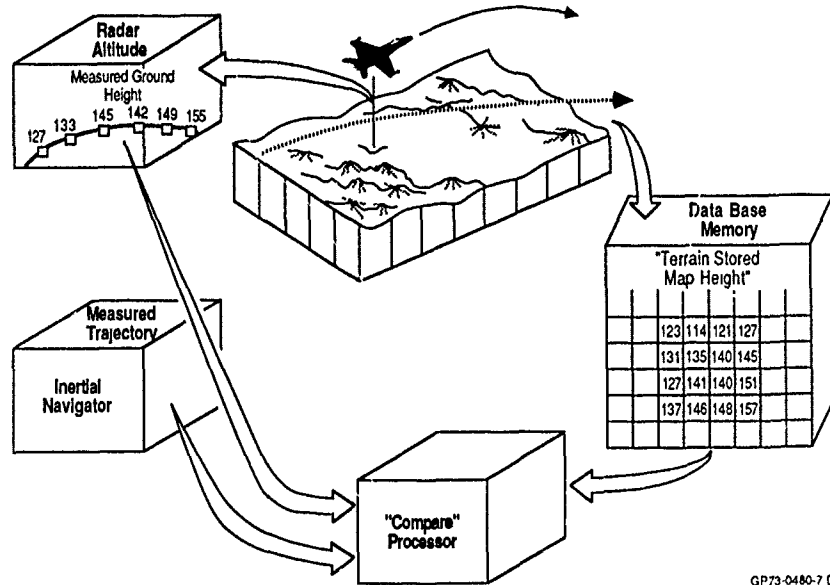
Global Positioning System (GPS) - The GPS is expected to become operational by 1990. GPS is a worldwide, satellite based navigation system which will provide highly accurate, jam resistant position, velocity and time to passive users. The system consists of eighteen satellites, in 12-hour circular orbits.

Each satellite completes two revolutions of the earth during a 24 hour period and the configuration permits an instantaneous view of at least four satellites at most points on the earth's surface. The satellite positioning permits the user to receive phase-modulated, binary-coded, wide-spectrum ranging signals from each of the satellites in view and to accurately determine position, velocity and time based on knowledge of the orbital geometry. Stated RMS accuracy of GPS is 5 meters in position and 0.1 meters/sec in velocity.

When fully operational GPS will synergistic with all terrain-aided navigation systems. As envisioned the GPS will be coupled with the aircraft's inertial navigator to provide precision navigation to the let down point of the low level penetration. At this point the system will probably switch over to a terrain-aided mode with the GPS acting as part of the navigation redundancy management system. The reason for this switch, even though GPS accuracy appears to be better than terrain-aided modes, is that there will be a bias between GPS and the terrain data. During low level penetration, the relative position of the aircraft with respect to the terrain will be more important than absolute position.

As the GPS is used to survey the earth's surface, more precise terrain data will be available with a common datum (GPS). More accurate terrain data and smaller cells will, in turn, increase terrain-aided accuracy and eliminate GPS/terrain biases. Future precision navigation systems will probably consist of dual, high-accuracy INS's with GPS and terrain aiding for the ultimate in jam-proof, covert TF/TA/TA. As the aircraft approaches the target, a transfer alignment will be performed from the aircraft precision navigator to GPS/inertially-aided munitions for stand-off weapon delivery.

Terrain-Aided Navigation - Terrain-aided navigation systems are computational techniques imbedded in software rather than specialized hardware. As shown in Figure 16 all terrain-aided systems use measured ground height from an altimeter, a data base memory with terrain stored map height, an inertial navigation system and a data processor. The processor uses specially developed algorithms to compare measured height from the altimeter with stored height from the data base to determine the INS position and velocity errors. These errors are then removed to provide navigation accuracies on the order of 30 meters to 100 meters, depending on the terrain being overflown and the accuracy of the data base.



GP73-0480-7 D

Figure 16. Terrain-Aided Navigation

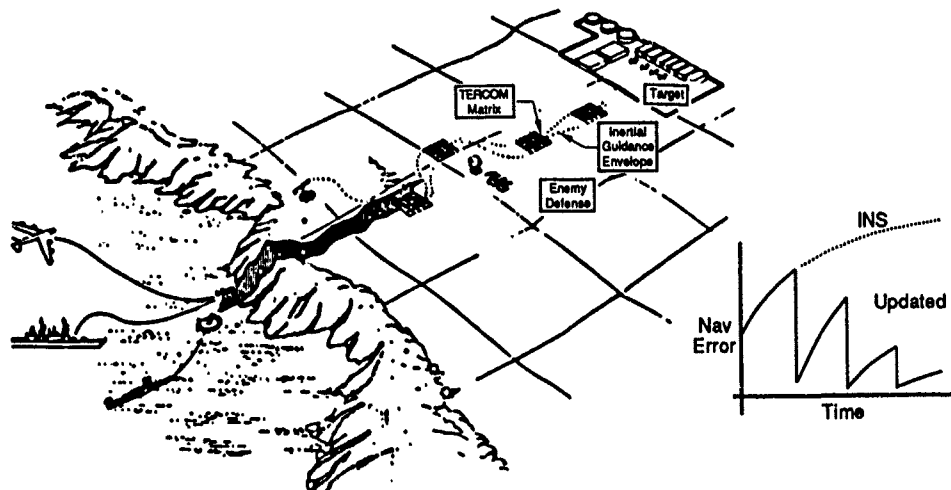
The benefits of terrain aided navigation in military applications are high-accuracy, drift-free navigation which gives a pilot a considerable edge. Good aircraft positional accuracy is vital for precision ground attack, particularly against high value, well defended targets. The attack can often be executed without the aircraft being exposed to enemy defenses since it will not have to fly over a visual identification point to initiate the attack. Moreover, it can perform an accurate transfer alignment to a stand-off weapon, such as an inertially aided munition, and thus only need to approach the target within several miles.

Another benefit is the ability of planning a mission without visual or radar-significant waypoints. Since terrain data is available and the aircraft position is well established within this datum, the terrain following ability of the aircraft can be enhanced. Even if the aircraft has a terrain following radar, information on terrain which is obscured and on obstructions is available. The aircraft can also approach the target covertly with no risk of being jammed, thereby complementing a forward looking infrared sensor or night vision goggles during night operations. A ground proximity warning system can also benefit significantly from a terrain-aided navigation system. To evaluate different approaches to terrain aided navigation, a literature survey was performed, the descriptions that follow are from that literature.

3.1 MANEUVERING TERRAIN CORRELATION SYSTEM: REFERENCE 2

A Maneuvering Terrain Correlation System (MTCS) has been developed which provides autonomous accurate positioning without requiring a visual overfly or a pop-up for sensor update. This system determines aircraft position by correlating a series of radar altimeter measurements of the height of the terrain below the aircraft with stored digital terrain data. The operating parameters of such a system have been defined and position detection method has been flight tested with promising results. The following summarizes study results and potential application to a penetrating tactical fighter aircraft.

Operational Concept - The classical terrain contouring matching (TERCOM) approach correlates altimeter measurements with stored digital terrain to determine position. In doing so, a straight horizontal profile is flown over a preplanned patch of terrain oriented parallel to the flight path. The patches are carefully preselected and pretested to ensure a high probability of detecting the correct position. Between patches, the vehicle relies on its INS for position information. Figure 17 illustrates the TERCOM concept.



GP73-0480-51-D

Figure 17. Cruise Missile Application of TERCOM

In contrast, the MTCS (Figure 18) achieves continuous correlation update, during maneuvering flights, throughout its area of operation using DLMS (Digital Land Mass System) data. The algorithm used for the digital correlation allows the aircraft the freedom to maneuver along highly curved, as well as straight, flight paths. Positional fixes are made continually and anywhere - even if the aircraft is flying over terrain with low terrain height standard deviation (σ_T).

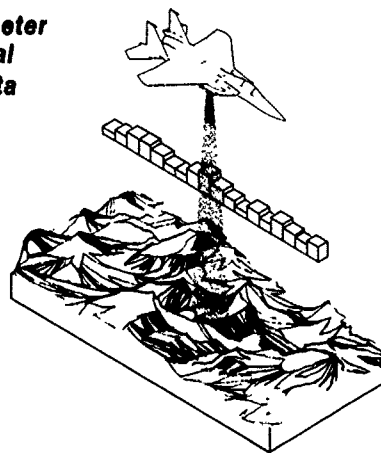
Correlates Radar Altimeter Readings with Digital Terrain Elevation Data

● **Hardware/Software Requirements**

- INS
- Radar Altimeter
- Stored DLMS Data
- Digital Correlation

● **Performance/Operational Features**

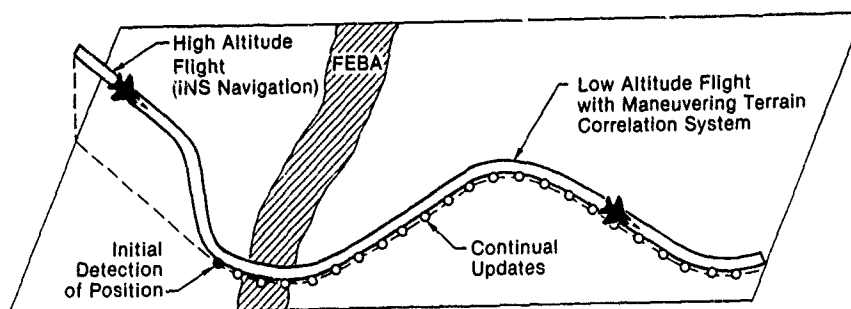
- Low Altitude
- All-Weather
- Autonomous
- Maneuvering Flight
- Continuous Update



GP73-0480-72 R

Figure 18. Maneuvering Terrain Correlation System

The operational concept for the system is shown in Figure 19. A tactical aircraft, after cruising at high altitude using its INS, develops an uncertainty in its position. Therefore, after letdown to penetration altitude, the MTCS must first detect its position. Following the initial detection, it provides continual updates of position. The system has to be capable of providing positional information even though the aircraft performs maneuvers for terrain following, terrain avoidance or threat avoidance. The system must be flexible enough that if a breaklock occurs, a redetection of position can be made. Breaklocks can result from a high bank angle (greater than the coverage of the radar altimeter) or from flying over large lakes or rivers.



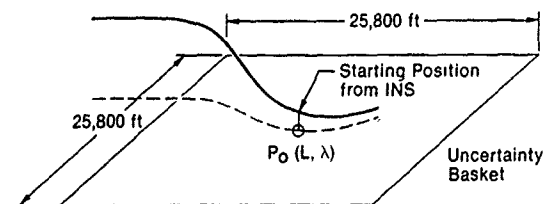
Technical Issues

- Initial Fix or "Capture"
- Curved Flight Path
- Operational Terrain Requirement

GP73-0480-73 R

Figure 19. Operational Concept

Initial Detection - The scheme for obtaining an initial detection of position is sketched in Figure 20. The aircraft has been flying at high altitude using its INS and has just let down to low altitude. The best estimate of the aircraft position (P_0) is the latitude and longitude coordinates given by the INS. However, the INS drift rate is known and boundaries can be placed around that estimated position to form an uncertainty basket in which the aircraft true position may be found. In the Figure 20 example it is assumed that the aircraft has been flying for approximately 45 minutes since its last update of the INS and that the INS has a drift rate of 1 nm/hr. After letdown the aircraft's radar altimeter begins making measurements of the height of the terrain below the aircraft and continues this until the aircraft has covered a specified distance called the "integration distance." Integration distances are kept small so that INS errors will not invalidate the recording of the relative positions between altimeter measurements. These altimeter measurements of ground height along the flight path are then correlated with ground heights from the digital terrain data for the uncertainty basket to find where the best match between data sets exists. Such a match gives the most probable flight path over the area and the most probable position where altimeter measurements were started. In order to avoid problems where a "false fix" could be obtained, at least M such probable positions must agree after N trials. That is, rather than use just one "most probable" position, a number (N) of these probable positions are obtained sequentially and at the end of all that data gathering, at least a portion (M) of the N probable positions must agree as to where the aircraft currently is.



1. Obtain Starting Position P_0 from INS (Lat, Long.)
2. Form Uncertainty Basket Around P_0 of $\pm 3\sigma_N$ Where.
 $\sigma_N = 1 \text{ NM/hr} \cdot \text{Number of Hours Since Takeoff or Update}$
 Design Point: $6\sigma_N = 25,800 \text{ ft} \times 25,800 \text{ ft}$
 $= 86 \times 86 \text{ Terrain Data Points} = 7,396$
3. Take Altimeter Readings of Ground Heights for Integration Distance
4. Correlate with Terrain Data
5. Require that at Least M Correlated Positions Agree

GP73-0480-74 R

Figure 20. Process for Initial Detection of Position

Figure 21 illustrates how the comparison of altimeter measurements and digital terrain data is made. An uncertainty basket of possible aircraft positions has been drawn around the estimated position. Every point in this uncertainty basket will sequentially be assumed as the starting point for the altimeter measurements. Using the matrix of incremental positions along the flight path, values of ground heights are found

along an assumed flight path over that portion of the digital terrain data. Thus, for each assumed starting position in the uncertainty basket a matrix of digital terrain ground heights is found. That matrix is subtracted from the matrix of ground heights sensed by altimeter as indicated by the equation. After subtraction, the absolute values of the quantities in the resulting matrix are averaged to produce the mean absolute difference (MAD). Mean values of the altimeter measurements and the digital terrain data are removed from the MAD value to eliminate the effect of bias errors in INS altitude above sea level and differences between datum levels in the digital data and sea level. After the entire uncertainty basket has been spanned and a complete matrix of MAD values formed, the MAD matrix is searched for the minimum MAD value; thereby finding the most probable position.

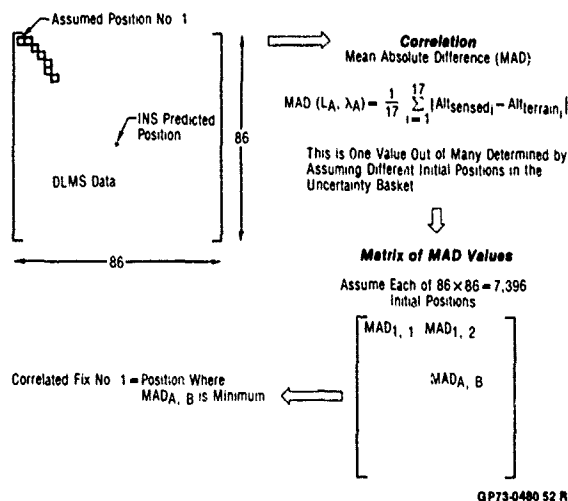
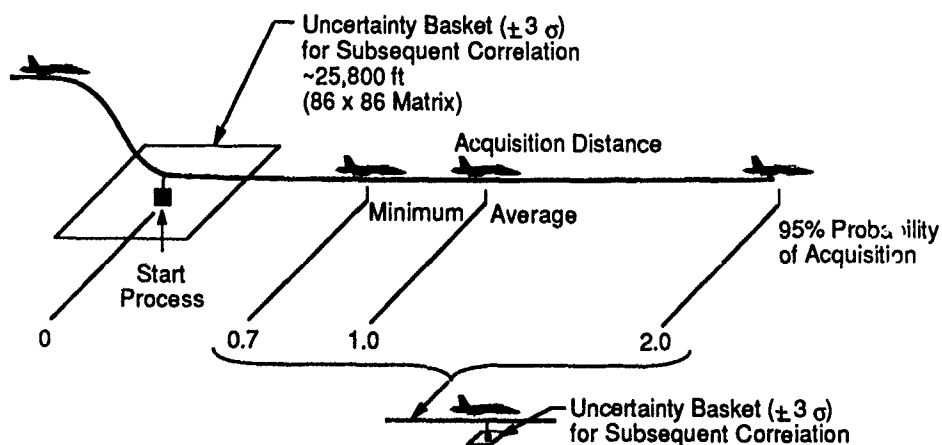


Figure 21. Initial Correlated Position

Operational Features - A summary of the complete detection process is shown in Figure 22. After letdown the correlation process begins with a large uncertainty basket. As the aircraft travels along, the earliest chance it has for detecting its true position is when it crosses the minimum distance line. Based on a large number of computer runs simulating the process, an aircraft will detect its true position after flying the average acquisition distance. After approximately twice this distance, the aircraft has a 95% probability that it will have determined its true position. After a true position is established, the uncertainty basket is reduced for subsequent correlations and the correlation process is continued to obtain better and better accuracy on the true position.



GP73-0480-57-D

Figure 22. Detection Process Summary
Normalized to Average Acquisition Distance

In terrains that are more rugged than that used in the simulations, the probability of detection increases and the average acquisition distances approach the minimum distance. The time between letdown and when the aircraft crosses the average acquisition distance line is shown in Figure 23 for various terrain roughnesses. Here the terrain measurement noise of 30 ft and a 4 out of 9 detection scheme was used. The rapid increase in acquisition time for smoother terrains is indicative of the effect of signal to noise ratio. The effect of terrain roughness on steady-state accuracy is also shown in Figure 23. Rougher terrains contain a greater signal content so accuracies improve. In general, the error in this first detection of true position is tolerable for the terrain of central Europe.

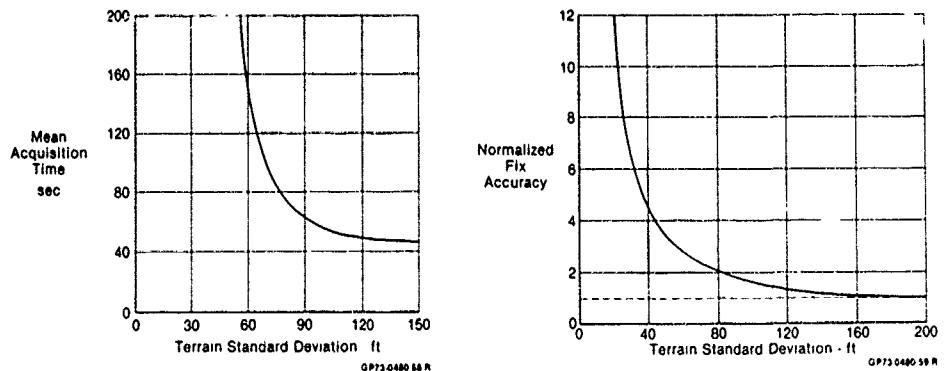


Figure 23. TERCOM Acquisition Time and Navigation Accuracy

Continuation of the correlation process after the initial detection improves the system accuracy. An example is shown in Figure 24 of average positional accuracy as a function of time. The accuracy of the true positions improves by approximately $1/N$ where N is the number of correlated true positions obtained.

In summary, the MTCS provides an autonomous, all weather means of accurately fixing aircraft position. It correlates digital terrain data with radar altimeter data in maneuvering flight. It has a low probability of false fix even over relatively smooth terrains. It uses current hardware but needs a fast mass storage system to realize its full potential. It is a self contained, highly accurate means of fixing position horizontally and vertically. We see it as an automatic update system which is compatible with advanced terrain following/terrain avoidance concepts.

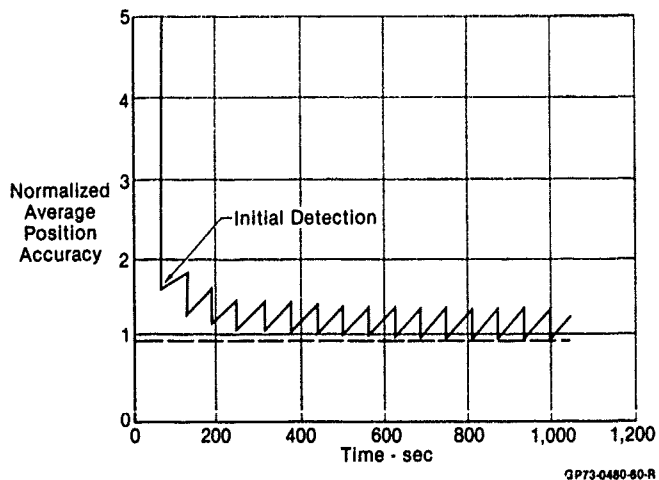


Figure 24. TERCOM Position Accuracy vs Time

3.2 SANDIA INERTIAL TERRAIN-AIDED NAVIGATION (SITAN): REFERENCE 3

SITAN is a computer algorithm that produces a very accurate trajectory for low-flying, high-performance aircraft by combining outputs from a radar or laser altimeter, an inertial navigation system (INS), and a digital terrain elevation map. An extended Kalman filter algorithm (Figure 25) processes ground clearance measurements continuously to estimate errors in a trajectory produced by an unaided INS. The filter is formulated with the following state model:

$$\underline{\delta x}_{k+1} = \Phi \underline{\delta x}_k + \underline{w}_k \quad (1)$$

and the measurement,

$$\begin{aligned} c_k &= c(\underline{x}_k) + v_k \\ &= z_k - h(\cdot, \cdot) + v_k, \end{aligned} \quad (2)$$

where

$\underline{\delta x}_k$ = INS error states to be estimated
 Φ = state-transition matrix for INS errors
 \underline{x}_k = states of INS and aircraft
 c_k = ground clearance measurement
 z_k = altitude of aircraft
 h = height of terrain at position (\cdot, \cdot)
 \underline{w}_k = driving noise with $E(\underline{w}_k) = \underline{0}$ for all k and $E(\underline{w}_k \underline{w}_j^T) = Q_k \delta_{kj}$
 v_k = measurement error with $E(v_k) = 0$ for all k and $E(v_k v_j) = R_k \delta_{kj}$
 k = subscript denoting time k .

1. Specify $\delta x_0^{(+)}$, $P_0^{(-)}$, Q , R , and Φ
2. Predict (Propagate) Error State and Covariance

$$\delta x_{k+1}^{(-)} = \Phi \delta x_k^{(-)}$$

$$P_{k+1}^{(-)} = \Phi P_k^{(-)} \Phi^T + Q$$
3. Use $\delta x_{k+1}^{(-)}$, $P_{k+1}^{(-)}$, and DTED to Compute H_{k+1} , $RFIT_{k+1}$, and $h(x_{k+1}^{(-)})$
4. Gain Computation

$$K_{k+1} = P_{k+1}^{(-)} H_{k+1}^T (H_{k+1}^{(-)} H_{k+1}^{(-)T} + R + RFIT_{k+1})^{-1}$$
 (Inverse is a Reciprocal in This Case)
5. Update the Error State and Covariance

$$\delta x_{k+1}^{(+)} = \delta x_{k+1}^{(-)} + K_{k+1} (C_{k+1} - (Z_{k+1}^{(-)} - h(x_{k+1}^{(-)})))$$
 For Acquisition Filters

$$P_{k+1}^{(+)} = (I - K_{k+1} H_{k+1}^{(-)}) P_{k+1}^{(-)}$$
 For Track Filter

$$P_{k+1}^{(-)} = (I - K_{k+1} H_{k+1}^{(-)}) P_{k+1}^{(+)} (I - K_{k+1} H_{k+1}^{(-)T}) + K_{k+1} (R + RFIT_{k+1}) K_{k+1}^T$$
6. Update Vehicle State Estimate

$$x_{k+1}^{(+)} = x_{k+1}^{INS} + \delta x_{k+1}^{(+)}$$
7. Return to Step 2.

OP73-0480-48 R

Figure 25. AFTI/SITAN Extended Kalman Filter Algorithm

Since the measurement function $c(\underline{x})$ is a nonlinear function of the states, the standard extended Kalman filter approach is used to obtain the measurement matrix,

$$H_k = \left. \frac{\partial c(\underline{x})}{\partial \underline{x}} \right|_{\underline{x} = \underline{x}_k^{(-)}} = [-h_x, -h_y, 1, 0, 0, \dots], \quad (3)$$

where h_x and h_y are the terrain slopes in the x and y directions of the map evaluated at $\underline{x}_k^{(-)}$, the predicted aircraft position just before a measurement is processed at time k . The first three states are taken to be the x position, y position, and altitude, respectively. At any time k ,

$$\underline{x} = \underline{x}_{INS} + \underline{\delta x}. \quad (4)$$

The state vector often used in a single filter implementation is

$$\underline{\delta x} = [\delta x \ \delta y \ \delta z \ \delta v_x \ \delta v_y]^T, \quad (5)$$

where δx , δy , δz , δv_x , and δv_y are errors in the x position, y position, altitude, x velocity, and y velocity, respectively. Other INS errors and states can also be included in $\underline{\delta x}$ by using the proper Φ .

Because the distance needed by SITAN to reach steady state becomes excessive as initial position errors approach several hundred meters parallel SITAN was developed. Parallel SITAN is a bank of extended Kalman filters that process identical altimeter measurements. After some updates, the filter with the minimum average weighted residual squared (AWRS) value is identified as having the correct position estimate. The AWRS value is defined by:

$$\text{AWRS}_{j\text{th filter}} = \frac{1}{n} \left[\sum_{i=1}^n \frac{\Delta_i^2}{H_i P_i H_i^T + R_i} \right]_{j\text{th filter}} \quad (6)$$

where Δ_i is the residual of the j th filter at the i th update, n is the number of updates, and $H_i P_i H_i^T + R_i$ is the residual variance. Once the initial position errors are reduced by the parallel filters, a single filter performs well, starting off essentially in steady state.

To implement parallel SITAN, the number and geometrical layout of the parallel filters needed to cover an initial position error must be specified. A square, constant-spaced grid can be used to center the filters about the horizontal position indicated by the INS. Filters at and near the corners are then eliminated to reduce the number of filters. To further lighten the computational burden, three-state, instead of five-state, filters are often used in parallel SITAN with

$$\underline{\delta x} = [\delta x \ \delta y \ \delta z]^T \quad (7)$$

For both the single and parallel filter implementation, a least-squares plane fit to the map, known as stochastic linearization, is used to compute the slope, h_x and h_y . Horizontal uncertainties σ_x and σ_y from the error-covariance matrix, defined by:

$$P = E[(\underline{\delta x} - \underline{\delta \hat{x}})(\underline{\delta x} - \underline{\delta \hat{x}})^T] \quad (8)$$

$$\text{Diag } P = [\sigma_x^2 \ \sigma_y^2 \ \sigma_z^2 \ \sigma_{vx}^2 \ \sigma_{vy}^2] \quad (9)$$

are used to determine the size of the plane. Residuals from the plane fit, R_{FIT_k} , are added to the measurement error variance, R_k , to ensure that the SITAN filter assigns less weight to the measurement when the plane fit is either very large or is over a rough area, thus adapting to local terrain.

On startup, SITAN enters an acquisition mode where 57 parallel, three-state filters are used to estimate the initial position error. During acquisition, position estimates are not used. When one of the 57 filters is identified as having a reliable estimate of the true aircraft position, a track mode is entered where a single, five-state filter is initialized at the selected acquisition filter's estimated position. During track, estimates of the aircraft's position are output every 100 m. The mode-control logic described later in this section handles the transition from one mode to the other.

Acquisition Mode - Acquisition mode is used to reliably and efficiently locate the aircraft's position within a large uncertainty region. The initial position error is covered with 57 three-state filters centered on a grid whose initial position estimates are 525 m apart. If the initial position error is described by a bivariate normal distribution with an initial CEP of 926-m, the probability of having the aircraft's position fall within the filter basket is 0.9785.

Lost Mode - If the mode-control logic determines during acquisition that the aircraft is not within the 2363-m radius search area, a lost mode is entered. When in the lost mode, SITAN does not provide any position estimates for the aircraft and the pilot has to update the INS to restart SITAN.

Track Mode - The purpose of the track mode is to continuously estimate the position of the aircraft as accurately as possible. The position estimate must be accurate since the aircraft computer uses the filtered SITAN-estimated positions to correct navigation errors.

Mode-Control Logic - With the design for the acquisition, lost, and track modes as described above, mode-control logic is needed to determine in which mode the algorithm should operate. When large aircraft position errors exist, it chooses the acquisition mode; with small errors, the track. The main parameter used in the mode-control logic for transition from acquisition to track is the AWRS.

Real-Time Implementation - The SITAN algorithm has been implemented on a single 28001 microprocessor running at 4-MHz. The SITAN module contained the acquisition mode, lost mode, track mode, and mode-control logic. The algorithm was implemented in assembly language using fixed-point calculations with a 3 Hz iteration rate. The acquisition Kalman filters were implemented with 16-bit word length variables, using the conventional form of the covariance update. The track filter is implemented using 32-bit word length variables and the stabilized "Joseph's" form of the covariance update. The worst case processor loading occurs in acquisition where the processor is 51% utilized. The remaining 49% is utilized by I/O conversions, data-base management, and operating system overhead.

Conclusion - The design of a recursive, computationally efficient, terrain-aided navigation system for aircraft has been implemented in a single Z8001 microprocessor. It gives median track accuracies less than 100 m over gently rolling, forested terrain using planar, cartographic-based terrain data. The results demonstrate that terrain data can be used in conjunction with the radar altimeter and INS that are normally present on an attack aircraft to produce trajectories whose accuracy is much greater than that available from the INS alone.

3.3 TERRAIN PROFILE MATCHING (TERPROM): REFERENCE 4

TERPROM combines the best features of the maneuvering terrain correlation system (MTCS) and the Sandia Inertial Terrain Aided Navigation (SITAN). Figure 26 illustrates the basic TERPROM principle, determining terrain height by differencing real time altitude above the terrain determined by an altimeter from flight path height above sea level.

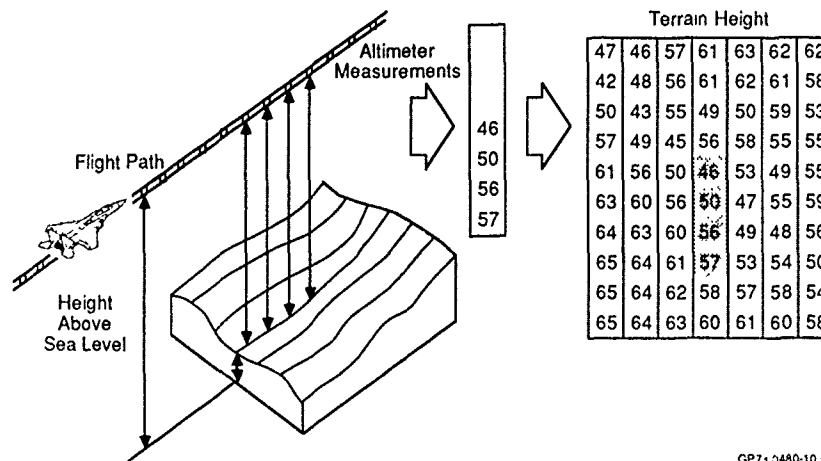


Figure 26. TERPROM Principle

GP73 0480-10 D

As the aircraft passes over the mapped area the computer samples the inertial instruments and corrects them for known errors. At each sample the indicated plan position is used to derive a ground height from the stored map and this, together with the indicated absolute height of the aircraft (again obtained from the inertial instruments and corrected for known errors), is used to obtain a prediction of the radar altimeter reading. This is compared with the actual radar altimeter measurement and the difference is processed by a Kalman filter to generate estimates of the inertial instrument errors which are then used to correct the states (position, velocity, etc.) of the system. This process, illustrated in Figure 27, is repeated every time the instruments are sampled, providing essentially continuous updating of the aircraft's position.

One of the key features incorporated into TERPROM is the ability of its Kalman filter to calibrate the inertial system and to propagate initial INS alignment errors.

Operating Modes - TERPROM takes advantage of the best features of MTCS and SITAN. It has three distinctive modes (1) acquisition mode, (2) continuous mode and (3) no update mode. Figure 28 illustrates the operating modes functional block diagram.

In the **acquisition mode** the Single Fix algorithm, rather than processing each altimeter measurements alone, constructs a profile of the ground over, typically, 5 kilometers. It compares this with profiles derived from its stored terrain model, the best fit determining the actual path taken. To derive the ground profile from the terrain model, the shape and orientation of the flight path is assumed to be that given by the inertial instruments; in other words, the unaided inertial system is assumed to have a constant offset for the duration of the TERPROM Single Fix run. Thus, the aircraft has complete freedom to maneuver, both vertically and horizontally, during the fix.

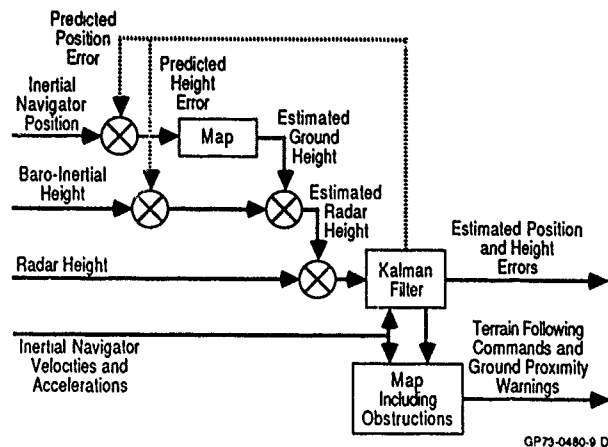


Figure 27. TERPROM System Flow Diagram

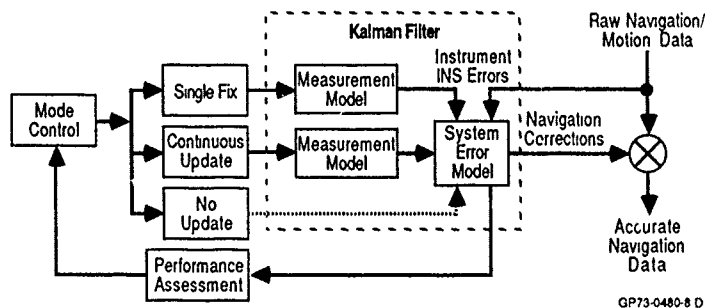
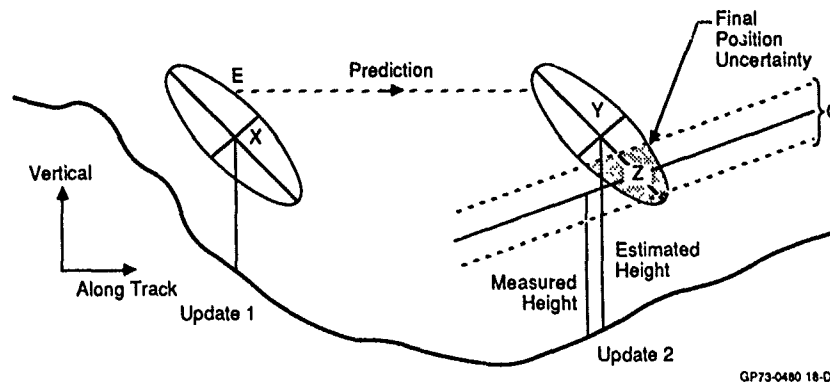


Figure 28. TERPROM Operating Modes

For any particular value assumed for the offset of the inertially derived position, a particular flight path is obtained and therefore a particular ground profile will be derived from the map. This profile is compared with that measured by the aircraft's radar altimeter and a 'figure of merit' for the match is computed. Different assumed offsets give different figures of merit, the true offset corresponding to the highest figure. Naturally, only discrete values of offset can be sampled in the test, but interpolation between these values is used to give a more accurate estimate of the true offset. The range of offsets considered must obviously be large enough to include the true offset. An estimate of the required range is computed from the uncertainty measure within the TERPROM Kalman filter which essentially predicts the propagation of errors from the alignment of the inertial instruments. In theory any range can be covered. But, in practice the number of sampled offsets is limited by the available computing power on the aircraft. A convenient limit suitable for most applications is about 1 nautical mile in any direction.

Occasionally another strip of ground has a very similar profile to the one overflowed resulting in a false fix. To ensure that this occurrence has no adverse effects on the system performance, a second Single Fix is performed and correlation between the two is sought. Only when correlation is found on the second or a subsequent fix is the Kalman filter initialized and the position information output.

An understanding of system operation in the continuous mode can be obtained by considering the simplified example illustrated in Figure 29 in which only along-track and altitude errors are considered. Having processed the altimeter measurement at update 1, the Kalman filter has an estimate of the aircraft's position and height (point X in Figure 29) and a measure of the uncertainty in its estimate. This is indicated by the error ellipse drawn at update 1 which will, in general, lie with its major axis parallel to the terrain surface at update 1. Between updates the Kalman filter propagates the position, height and their uncertainty information forward to update 2 using a dynamic error model of the inertial measurement unit. This predicted position is marked Y and its uncertainty region is illustrated by the propagated error ellipse. The altimeter measurement, at update 2, by itself would indicate an uncertainty region (G) parallel to the new terrain tangent. The Kalman filter optimally merges this new information into an improved position and height estimate, Z, having an uncertainty region smaller than that of either measurement taken separately. This has been conceptually indicated at update 2 by the double cross-hatched region. Note that successive improvements in accuracy are a function of the terrain slope, the slope variation and the measurement error.



GP73-0480 18-0

Figure 29. TERPROM'S Continuous Mode

At times, such as when passing over extensive stretches of water, the terrain has no profile to match and the system enters the no update mode. However, as the Kalman filter in the TERPROM System has estimated the errors in the inertial instruments, these instruments are seen as if they were more accurate than before; that is the INS has been calibrated by TERPROM. This factor, used with TERPROM's model of error propagation, means that accurate navigation can still take place when no usable terrain profile exists. On re-entering an area suitable for updating the Kalman filter, an estimate of the aircraft's likely position error is assessed to determine whether the Continuous Update can still be used, or whether a confirmatory Single Fix should be obtained. Similarly, if at any time the input data starts to behave in an unexpected manner, as would occur, for example, with gross map errors or radar altimeter jamming, TERPROM temporarily rejects the faulty data in order to maintain system integrity. Subsequently the Single Fix algorithm can be used to confirm or correct TERPROM's operation, as this algorithm is stable even in the presence of large position errors.

Of all the terrain aided aircraft navigation systems discussed, TERPROM is the most mature. Although MTCS and SITAN have been flown in aircraft, TERPROM has been under continuous development since 1975 and has a large number of development flights in fighter aircraft commencing in 1985.

3.4 SUMMARY

Three terrain aided navigation systems using radar altimeter measurements and a terrain height data base are available. In the late 1980's all have a similar position accuracy of 30-100 meters CEP. They all provide an autonomous, all weather means of accurately fixing aircraft position horizontally and vertically and are compatible with the survivable penetration concept. In the 1990's, when GPS becomes operational, a combined GPS- and terrain-aided/INS will become the navigation system of attack aircraft.

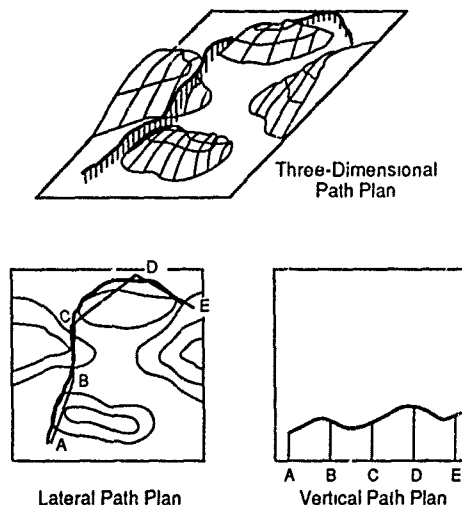
4. TERRAIN FOLLOWING/TERRAIN AVOIDANCE/THREAT AVOIDANCE

A penetrating aircraft attempts to achieve covertness to increase its survival probability and achieve tactical surprise as it delivers its weapons. Covertness is achieved by reducing the observability of the aircraft and by flying a path which minimizes detection by hostile sensors. Although much effort is expended to minimize the radio frequency, infrared, visible, and acoustic signature of an aircraft, the necessity of moving tens of thousands of pounds of metal through the air at supersonic speeds prevents the fielding of an invisible aircraft. The generation of a flight path which hampers detection and attack by elements of the hostile air defense system is a necessary component of covert penetration. Because of the density of air defense systems in modern environments, penetration flight paths which avoid the detection envelope of all elements of the air defense system are difficult, if not impossible to achieve. In high threat density environments, avoidance of the lethal envelope of all threats is also difficult to achieve. The generation of a flight path which minimizes detection by threat sensors and minimizes exposure to lethal threats is a crucial part of any covert penetration system.

In general, exposure to detection is decreased by flying as low as possible, consistent with flight safety. Flying at low above-ground-level (agl) altitudes allows the penetrating aircraft to be masked by terrain intervening between a hostile site and the flight path. When the avoidance of detection is impossible, low agl flight minimizes exposure time, thus increasing aircraft survivability. In recognition of these facts, terrain following systems have been built and fielded for many years. These systems have used a forward-looking radar and an evolving set of algorithms to permit safe flight at

altitudes on the order of 50 meters at speeds in excess of Mach 1. Mission planning included the definition of a route which sought to avoid known (and probable) threat locations, and provide good terrain masking in general. The terrain following algorithms were based on continuously collected radar measurements of the terrain ahead of the aircraft. Because of sensor limitations aircraft maneuverability was restricted; the flight path was essentially a sequence of straight-line segments between preplanned waypoints. Because the radar was the only source of terrain data, essentially continuous radiation, at relatively high power, was required. Technology is now available to develop systems which conduct terrain following, terrain avoidance, and threat avoidance more effectively. Digital data storage technology can now allow terrain information, at resolutions on the order of 100 meters (horizontally), to be carried on a penetrating aircraft for areas covering a tactical or strategic mission. Digital terrain data is available for much of the surface of the planet which is of military interest. Data processing technology has advanced to the point where sophisticated algorithms to define an effective three-dimensional flight path may be executed in real time, again within the space weight, and power constraints imposed by modern aircraft. Sensor technology will now permit intermittent operation with varying power levels, and allow rapid scanning over a wide angular volume.

The following discussion will focus on the algorithmic and system level aspects of covert flight path generation. The general problem of path planning in a three-dimensional space will be discussed in terms of two two-dimensional problems, the vertical and the lateral (See Figure 30). Although the solution technique used to attack the problem may operate explicitly in three dimensions, the two-dimensional approach is conceptually appealing and may be computationally attractive. This approach takes advantage of the large technology base in vertical control developed under prior terrain following systems and the "natural" perspective of lateral route planning in a plan view. The two problems are not independent, however. Data for vertical path generation must be obtained from the terrain over which the lateral path is drawn, and the lateral path generation algorithms must not generate paths that the aircraft will be unable to achieve in the vertical dimension. Additionally, the "natural" coordinate system for generating commands to the aircraft is in its roll, pitch and yaw axes; commands for a path generated in earth coordinates must be transferred into the aircraft coordinates.



GP73-0480-17-D

Figure 30. Two and Three Dimensional Solution Approaches

It is the availability of digital terrain elevation data (DTED) which allows feasible terrain following/terrain avoidance/threat avoidance systems to be built. Early terrain following systems processed each radar return as it was received, with no storage of the sensed terrain elevation. The LANTIRN terrain following radar system, now entering production, stores the data generated on a scan of the antenna and processes the stored data. Multiple scans at different azimuth angles are stored and this data used to generate vertical commands when the aircraft is turning. The data in LANTIRN is limited by radar range and line-of-sight restrictions--the radar cannot "see" through a hill. The availability of DTED can permit algorithms to "see" terrain masked from the radar and "see" terrain over which the radar has not scanned. This data can be used for lateral path planning, as well as for the generation of terrain following commands. The sensed data can be blended with the DTED so that commands can be generated from data representing the true terrain, as opposed to being based solely on prestored data. (The blending process can compensate for both errors in the DTED and misregistration between the true terrain and the digital map.)

4.1 DIGITAL MAP REQUIREMENTS

A source of digital terrain elevation data is an essential part of a terrain following/terrain avoidance/threat avoidance system using the concepts to be described. As shown in Figure 31, the digital map must contain data over the entire area encompassed by a mission. Because of tactical uncertainties, the area must be much larger than that immediately surrounding the waypoint-to-waypoint path must be stored. Areas of 30,000 to 100,000 square kilometers are not unreasonable for deep strike interdiction missions.



GP73-0480-50 R

Figure 31. Potential Digital Map Coverage

Data with a horizontal resolution of about 100 meters has been demonstrated to be effective for reasonable vertical and lateral path generation and for computing aircraft-threat intervisibility. At this resolution, some 10,000,000 terrain data points are required. A map covering the area from latitude 40 to 70 degrees north and longitude 10 to 40 degrees east would contain on the order of 1,000,000,000 points. Technology to support the mission requirements is currently available and that for the theater requirement will be available in the near future.

If data is not to be generated for each mission, a common coordinate system must be used for all maps. The United States Defense Mapping Agency produces a standard product of DTED in a latitude-longitude coordinate system with approximately 100 meter resolution. A latitude-longitude system is the natural system for aircraft navigation and mission planning. Data in such a system can readily be converted to any local coordinate system required by the path generation algorithms.

The vertical accuracy required of the DTED depends on the use to which it will be put. If a sensor is used to verify and correct digital map data at ranges within three to five kilometers of the aircraft, vertical errors of 20 to 30 meters may be tolerated. If no sensor is used, vertical errors must be guaranteed to be less than the agl altitude

of the flight path. DTED errors indicating terrain higher than the actual terrain cause the aircraft to fly higher agl than desired, increasing exposure and hence reducing survivability. DTED errors indicating terrain lower than the actual are catastrophic.

It is impossible to precisely (within 100 meters laterally and 5 meters vertically) predict the position of an aircraft throughout the course of a penetration mission extending hundreds of kilometers through hostile territory. The lateral and vertical path generation algorithms must thus be executed in real time, considering the current position and attitude of the aircraft, and the possibility of changes in the mission plan due to unforeseen threats, battle damage, or retargeting. It is computationally infeasible to compute a detailed path for the entire course of the mission in real time since the path generation algorithms have a finite look-ahead range. Because of these considerations, the algorithms must continually access the data in the digital map. The map must provide the bandwidth and response time to support the algorithms. Depending on the form of the algorithms, this may consist of data quantities ranging from a few hundred to tens of thousands of points per second.

4.2 TERRAIN AVOIDANCE/THREAT AVOIDANCE

The goal of a terrain avoidance/threat avoidance algorithm is to determine a lateral path such that the survivability of the penetrating aircraft is maximized. The path which the algorithm generates must follow the general route defined in the mission planning process. The mission plan and route (Figure 32) include a series of waypoints, spaced (nominally) tens of kilometers apart. Waypoints associated with the target or initial point of attack must be approached relatively precisely, while intermediate waypoints mark relatively gross changes in the desired route and need not be precisely overflown. Time on target is included in the mission plan, with tolerances varying from mission-to-mission (timing for multiple attacks with nuclear weapons on a target complex is much more critical than for an independent attack with conventional weapons). Timing and fuel management considerations dictate a maximum lateral deviation from the waypoint-to-waypoint path; the route consists of a pre-defined corridor. The route defined by the mission plan is designed to take advantage of terrain masking from known threats, and provide a relatively "safe" corridor for ingress and egress. The terrain and at coarser resolution out to a distance of less than 40 or 50 kilometers. The generated path is determined by two factors: the techniques used to generate candidate paths and the function used to evaluate the paths.

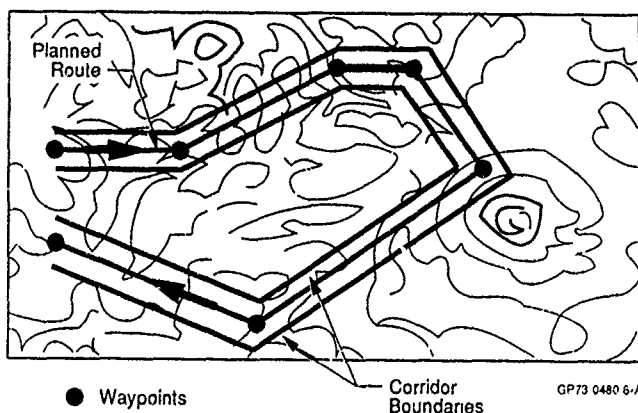
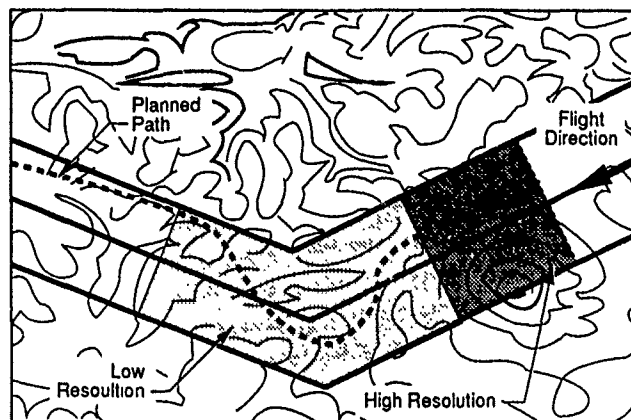


Figure 32. Mission Route

As illustrated in Figure 33, it is unnecessary to have the lateral path planned to an accuracy of better than 100 meters for a distance ahead of the aircraft of more than 2.5 to 5.0 kilometers. A pilot cannot be expected to follow a path that precisely; moreover, even if he (or an autopilot) could, the tactical environment may dictate deviations for reasons not comprehended by the algorithm. Path planning on a coarser scale than 100 meters, but finer than the corridor width, is necessary out to distances of 20 to 30 kilometers to position the aircraft to fly on the proper side of a hill or select the proper valley. Planning for shorter distances may make short, deep valleys attractive at the expense of increased exposure when leaving the valley.



GP73-0480-4 A

Figure 33. Resolution Variation in Route Planning

The basic function of a terrain avoidance/threat avoidance algorithm can thus be summarized as an attempt to generate a flyable path toward a predefined waypoint, within the constraints of a specified corridor, at high resolution for the next few kilometers and at coarser resolution out to a distance of less than 40 or 50 kilometers. The generated path is determined by two factors: the techniques used to generate candidate paths and the function used to evaluate the paths.

Path Cost Functions - The path generation problem may be cast as an optimization problem, finding a lateral path that optimizes some measure of performance. Since, in general, the objective is to fly low and avoid hostile air defenses, a minimization process is suggested wherein the best path is the one with the lowest cost. The cost function is used to evaluate the desirability of alternate paths. The cost function must be defined for each point that the path generation algorithm will evaluate for inclusion in a path. The cost function must operate to avoid threats, or at least minimize exposure in areas where avoidance of all threats is not possible. The cost function, in the absence of specific threat information, should serve to fly the aircraft over rugged terrain, thus providing masking against unknown threats. The cost function, however, should not cause the aircraft to fly over the lowest portions of a valley, since that is where the lines of communication normally are present. Valley flights should be above the floor, but below the ridgeline. The cost function should contain some provision for altering the relative weighting between terrain and threat costs. The cost function should contain some penalty for maneuvering, for crew comfort, fuel conservation and mission timeline compliance.

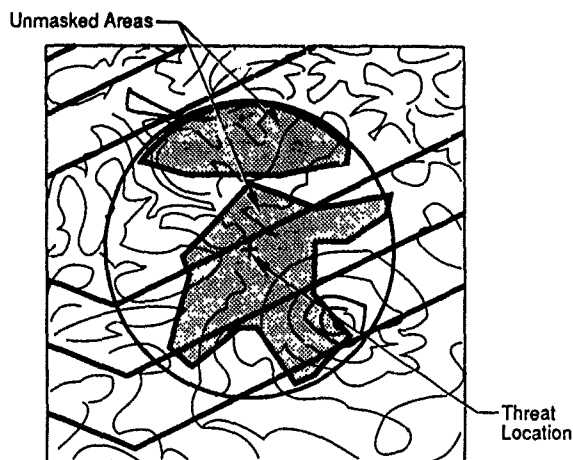
Cost functions may be constructed in a wide variety of ways. Cost functions which are additive, accumulating cost over the flight path are intuitively appealing and mathematically tractable. The cost function may thus be written as:

$$\text{PATH COST} = \text{SUM OVER THE PATH } \{ (P) * (\text{TERRAIN COST}) + \\ (1-P) * (\text{THREAT COST}) + (\text{MANEUVER COST}) \}$$

where P is a parameter ranging from 0 to 1 to adjust the relative weighting of terrain and threats.

Terrain cost is conveniently expressed as altitude above mean sea level of the terrain overflown by the path. Local normalization of the dynamic range of the terrain (considering terrain maxima and minima and variability over the area being considered) permits the same qualitative behavior of the flight path over rough or smooth areas. Taking the terrain cost to be absolute value of the difference between the terrain elevation of a cell and the mean value over a local area biases the generated paths away from the valley floors. The maneuver cost may be scaled to allow any desired degree of maneuverability in the path. The definition of threat costs is more complex.

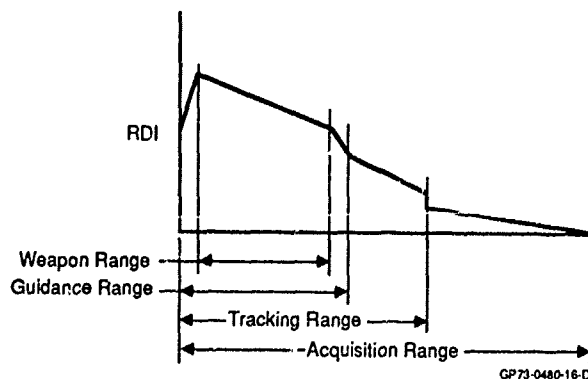
A hostile air defense site is a threat to a penetrating aircraft if there is clear line-of-sight between the aircraft and the site. The first task in determining a threat cost is thus the determination of the areas where the aircraft will be unmasked from the threat. This determination consists of examining the terrain between the threat site and the aircraft to determine if there is any intervening terrain. This may be done for the range of agl altitudes that the aircraft is anticipated to fly. These unmasked areas may be stored point-by-point, in the same data structures as are used for terrain data storage (to facilitate later processing), or may be stored by approximating the unmasked area by a polygon (See Figure 34) and storing the vertices of the polygons. Polygonal storage requires less memory space than does the marking of unmasked terrain, at the expense of computations necessary to determine if a given terrain cell is in the interior of any polygon. In either event, once the unmasked terrain is identified, it is necessary to estimate the lethality of the threat.



GP73-0480 5-A

Figure 34. Threat Polygon

The dangerous area around hostile air defense systems may be characterized by three radii; the acquisition range, the tracking range, and the weapon guidance range. Lethal damage to the aircraft may occur only within the weapon guidance range. Flight within the acquisition or tracking range, but outside the guidance range, does pose a danger to the aircraft since acquisition and track data from a site may be passed to other sites through the air defense network. As shown in Figure 35, a relative danger index (RDI), as a function of the threat parameters, may be computed within the unmasked areas within each of the radii. This index is higher in the lethal volume and typically decreases with range from the site. Taking the total RDI for any terrain cell as the sum of the RDI's for all threats which are visible from the cell serves to provide a means for accounting for the overlapping fields of fire often found in dense threat environments. As with the terrain costs, an "automatic gain control" function which normalizes the threat costs allows consistent algorithmic behavior in dense and sparse threat environments.



GP73-0480-16-D

Figure 35. RDI Variation

If the threat site location is known imprecisely (Figure 36), the threat location can be taken as on the highest point of terrain within the uncertainty area and the threat radii increased by the uncertainty. This generally provides maximal unmasked area and yields a conservative estimate of "safe" terrain. It is preferred to postulate that the aircraft is unmasked, when it in fact is masked, than to fly through a presumably "safe" area which is in reality exposed to hostile fire. Where the uncertainty in threat position is large compared to its lethal radius or where the threat is known to be mobile (and hence cannot be a-priori located precisely), a different technique for estimating RDI must be used. Masking is more effective in rough terrain than in smooth terrain. For areas where a threat is thought to operate but located very imprecisely, it is better to fly through the roughest part of the area than the smoothest. Terrain roughness may be estimated from the terrain data and rough areas assigned lower RDI values than smoother areas. Again, fundamental threat lethalties should be used to scale the RDI.

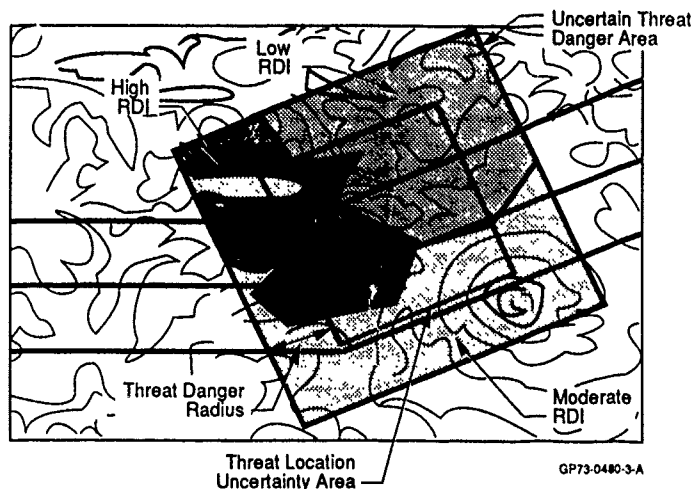


Figure 36. RDI for Threats With Large Location Uncertainty

The cost functions, terrain, threat, maneuverability, and possibly other terms provide the objective function which the path finding algorithm optimizes. Although the cost functions are quantitative, there is a certain degree of arbitrariness and uncertainty included in them. Cost functions of the type described above are used in lieu of a detailed estimation of the aircraft survival probability along all possible paths throughout the whole mission. Although survival probability is monotonically related to these costs, the correspondence is probably not one-to-one or linear. These considerations allow some freedom for the design of the optimization algorithms.

Lateral Path Generation - The data that the lateral path generation algorithm has to operate with are terrain elevation and threat relative danger indices (RDI) for the area over which the aircraft may fly, and the dynamics (and dynamical limits) of the aircraft. The terrain data is supplied in latitude-longitude coordinates; the RDI data may be available in these or other coordinates. Data spaced equally in latitude-longitude coordinates is nonuniform; the East-West distance between points is a function of latitude. A path generation algorithm using this data must either comprehend the variation in point spacing, or remap the data to a local coordinate system with the spacing used by the algorithm. For relatively small areas (on the order of 100 by 100 kilometers), a Universal Transverse Mercator (UTM) coordinate system may be used. Discontinuities at the transition from one standard UTM grid to another must be dealt with. If the operational area is small enough, a nonstandard UTM system may be used to construct a regular grid of points. The discrepancy between true North and the direction of the "columns" of the UTM processing grid must be accounted for in any case.

Since the path generation algorithms have a limited look-ahead (tens of kilometers as compared to a mission length of possibly hundreds of kilometers), it is advantageous to operate the algorithms in a Local Area Map (LAM), as illustrated in Figure 37. The use of a LAM allows the data to be structured so that the algorithm may execute efficiently and minimizes the storage requirements for immediately available data. The LAM may also be oriented for efficient algorithmic execution. An algorithm which reorients the LAM at every cycle of execution must perform the coordinate transformation from latitude-longitude to LAM coordinates each cycle. The reorientation may, however, allow the algorithm to execute significantly more efficiently. An alternate to reorientation at each execution cycle is to incrementally advance the LAM in the direction of aircraft movement, performing the transformation from latitude-longitude coordinates to LAM coordinates only for the added data. Incremental accessing of the DTED from the digital map is possible in either case, with the reorientation process requiring more computation and/or more local storage.

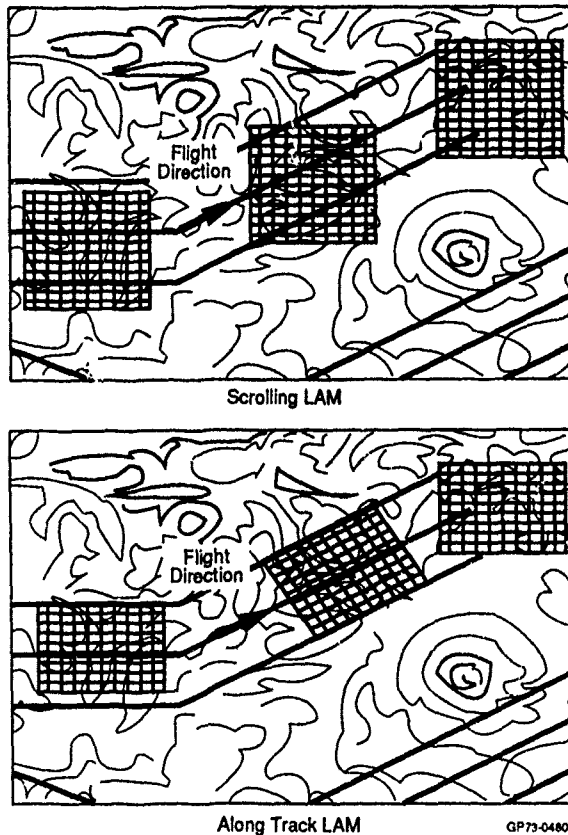
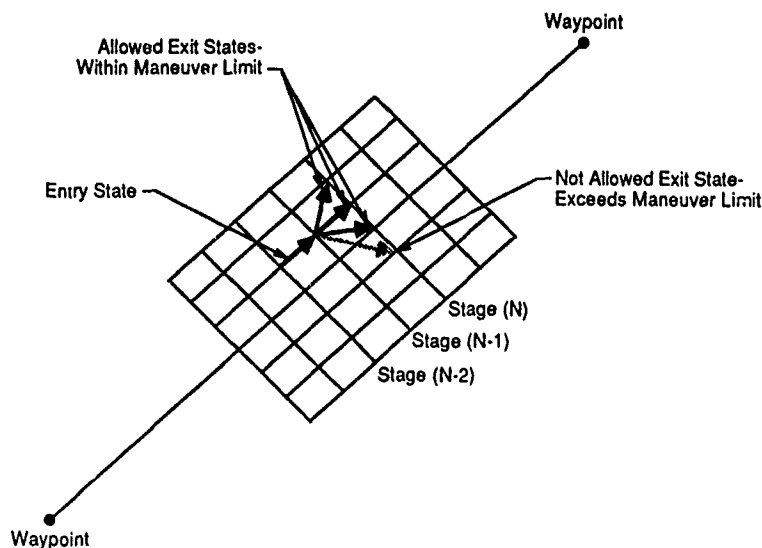


Figure 37. Different LAM Movement Approached

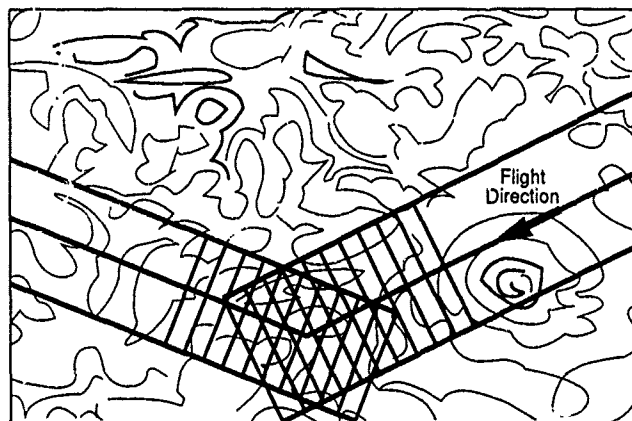
As has been described, the path generation algorithm must solve an optimization problem consisting of a linear objective function and a set of nonlinear constraints. The nonlinearity in the constraints arises because of the differential equations describing the aircraft dynamics. A straightforward definition of the algorithm may be obtained by explicitly writing the objective function and the aircraft dynamical equations, requiring that the aircraft maintain an agl altitude no less than a specified set clearance level. Mathematical programming techniques are available to solve this problem and have been applied successfully. This approach, because of the complexity of the problem and the iterative nature of the algorithms which must be used, results in a computational burden which is not consistent with real time operation in current tactical computer capability.

The geometric nature of the problem, wherein the aircraft must fly from point to point making progress along the general direction to the next waypoint along the route developed in the mission plan, suggests the use of dynamic programming as a solution technique (Figure 38). If a grid of points is oriented along and perpendicular to the waypoint-to-waypoint line, a dynamic programming formulation which uses the perpendicular rows of points as stages is possible. From any point in a given stage (row), the points in the next stage (row) which can be reached are limited by the aircraft dynamics. A reasonable approximation to the detailed dynamics allows the stage-to-stage transitions to be analytically described and easily mechanized. If the grid spacing is fairly coarse (on the order of between 500 and 1500 meters), the number of stages and points which must be examined for each transition allows execution of the algorithm in the time the aircraft can overfly one of the grid cells. "Turning the corner," changing direction when passing a waypoint, requires some additional computation. As shown in Figure 39, the process must be executed for each segment of the route within the current algorithm window. The paths for the consecutive two segments must then be connected while maintaining the lateral maneuver limits of the aircraft and achieving the specified tolerance on approach to the waypoint. The slope of the terrain between successive path points may be computed to prevent the generation of unflyable paths for climb-limited aircraft. The symmetries available in a triangularly spaced (rather than rectangularly spaced) data can be used to simplify the algorithm, at the expense of more complex maintenance of a minimal-memory LAM.



GP73-0480-15-D

Figure 38. Maneuver Constraint Enforcement in Dynamic Programming



GP73-0480-1 A

Figure 39. Dynamic Programming Corner Turning

Grid sizes of 500 to 1500 meters are inadequate for the high resolution processing required within the first two to five kilometers ahead of the aircraft. The DTED is available with three arc second spacing in latitude and longitude; this corresponds to about 100 meter spacing on the equator. East-West spacing decreases with latitude, until the available resolution changes to six arc seconds in longitude at 50 degrees latitude. LAM resolution of 100 meters may be achieved by interpolating the latitude-longitude data. The algorithm for generating this portion of the path need not be the same as that used for the farther distances.

Although dynamic programming may be used for the near-aircraft portion of the path, the necessity to explicitly consider the aircraft's state and dynamics negate much of the simplicity of the dynamic programming algorithm. Figure 40 illustrates an approach which has been successfully applied consists of generating a tree of potential paths, limited by the bounds of aircraft performance, and including explicitly the changes in aircraft state necessary not only to fly the lateral path, but also to maintain the vertical set clearance. Using the same data structures as for the dynamic programming LAM, but at the finer resolution, allows commonality of interface to the digital map and commonality of LAM management functions.

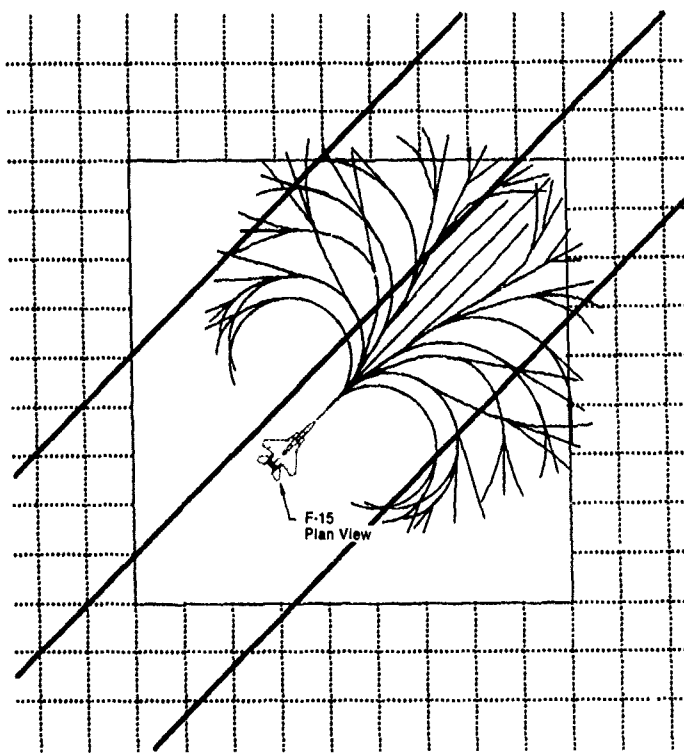


Figure 40. Close-In Tree Structure

4.3 TERRAIN FOLLOWING

Terrain following systems have been operational since the late 1960's. These systems have proven their tactical utility in combat. The currently fielded systems all use microwave radar as the terrain sensor. Systems have been postulated and experimented with using a variety of sensors, including unaugmented DTED. The systems operate by generating commands to cause the aircraft to fly above the terrain without violating a specified agl altitude minimum. The algorithms used for systems with microwave radar sensor may be adapted for use with data derived from other sources.

Some terrain following systems (all of the early ones) generate commands on a pulse-to-pulse (or group of pulses) basis, passing to the pilot or autopilot the maximum command generated over a vertical scan of the radar. The command generation algorithm uses a model of the aircraft dynamics, including maximum vertical acceleration and climb limit capability, to derive the command necessary to allow clearance of the measured terrain by the set amount. Monopulse techniques are used to reduce the elevation uncertainty of the measurements, with measurement errors of only a few meters at ranges less than five kilometers. Generally, only one sample per pulse repetition interval (the range cell containing the return from boresight) is processed, although work is being done with off-boresight processing techniques. The commands are presented to the pilot via a command indicator, or displayed on the Head-Up Display, and in many systems coupled to the autopilot for automatic terrain following.

The terrain following radar system in the LANTIRN Navigation Pod uses stored radar measurement as the basis for command generation. Radar returns from a vertical scan are processed to produce a terrain range-elevation profile, referenced to the aircraft position at the time of scan (See Figure 41). This data is adjusted for aircraft movement and commands are generated from the adjusted data. As shown in Figure 42, this processing allows the radar to scan off the current flight path to collect data for use when the aircraft begins to turn. When a turn is detected, the radar scans into the direction of the turn, collecting data and storing it in azimuth bars. The command generation algorithms extract data from the bars along a predicted flight path.

The LANTIRN system uses ADLAT command generation algorithms which develop vertical accelerations designed to fly a sequence of parabolic paths (Figure 43). These paths not only maintain the desired set clearance agl, but cross over isolated peaks with level flight. Positive and negative accelerations are generally limited by ride quality settings. Flight segments between parabolas, or when the preset climb (or dive) angle limits are reached, are straight lines tangent to the parabolas.

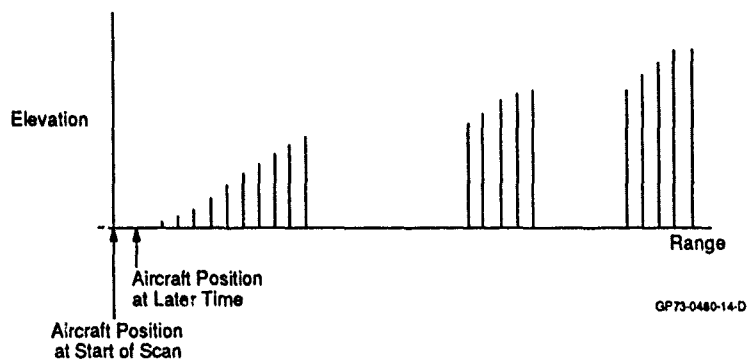


Figure 41. Range-Elevation Profile

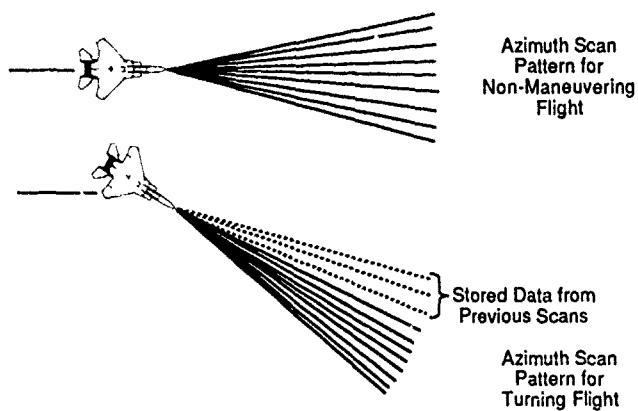


Figure 42. Azimuth Scan Patterns

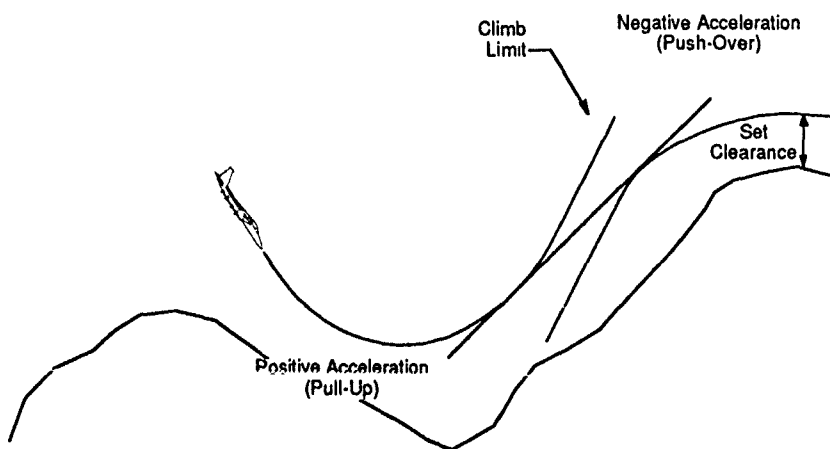


Figure 43. ADLAT Vertical Profile

The maximum range limits of the terrain following system are set by the necessity to see severe terrain discontinuities in time to execute smooth, safe clearance maneuvers. Ranges of 10 to 20 kilometers are adequate for aircraft with climb angle limits of about 10 to 15 degrees. Azimuthal beamwidths of about five degrees allow normal minor course changes to be accomplished within the data collected on one vertical scan.

While current terrain following algorithms were developed and are being used with radar sensed data, there is no intrinsic reason why data derived from other sources may not be also used. The Advanced Low Altitude Techniques (ADLAT) algorithm used in the LANTIRN system is particularly adaptable to using stored or blended data. These algorithms currently operate with stored data, buffered from the radar collection process by the scan bars. The command generation algorithms have no knowledge of the source of the data.

Current algorithms assume that the aircraft will maintain its current state. That is, if the aircraft is not turning, the algorithms use data directly ahead of the aircraft. If the aircraft is turning, a continuous turn at the current turn rate is assumed. The ability to collect and process data during turns places limits on the turn rates for which the systems can guarantee safe terrain following flight. Operation from a stored LAM containing a blend of DTED and radar-sensed terrain elevation data can remove some of the limitations of conventional system operation. If a lateral path generation algorithm is operating and the aircraft is being directed along a nonuniformly curving path, data may be extracted from along the path, organized into the format of the LANTIRN scan bars and presented to the ADLAT algorithm.

4.4 SENSOR/MAP DATA BLENDING

Because of the presence of errors in the DTED and the possibility of unmapped, man-made obstacles (deliberately placed along potential penetration paths by a hostile enemy), digital terrain data alone is insufficient to use as a basis for safe low level flight. Historically, microwave radar has been used to obtain the data for the generation of terrain following commands. Microwave, millimeter wave, or optical (laser) radars may be used to obtain terrain elevation data along the projected flight path. FLIR systems, with passive ranging capabilities may be feasible. The data from the aircraft sensors, whatever they may be, must be blended with the DTED to make use of all available information. Several strategies for data blending are possible.

One simple approach to data blending is to use the sensor data, where it is available, and map data elsewhere. The sensor can be directed to collect data over the projected flight path. The rationale behind this approach, particularly for microwave radar sensors, is that in the absence of the DTED the sensed data would be used to generate commands. This approach does not take advantage of the possibility that there may be obstacles which are mapped, but not seen by the sensor; sensor performance may be degraded by undetected failure; or an obstacle is obscured by environmental conditions.

A second simple approach to data blending is to use the maximum of the terrain elevations indicated by the DTED or the sensor. This approach is conservative in that mapped (but unseen) or seen (but unmapped) obstacles will be overflowed. The price of the conservatism is a "fly high" error in terrain following, causing the aircraft to have increased exposure to threats. In particular, if significant (man-made or natural) features are mislocated in the DTED disastrous pop-ups may be commanded.

A more operationally useful and sophisticated approach to data blending consists of weighting the sensor and DTED data according to some derived measure of confidence in each. Confidence in sensor data may be estimated by such parameters as received signal-to-noise ratios, image quality, or through a systematic re-examination with the sensor if discrepancies are noted between the DTED and sensor data. Estimates of DTED quality can be made by the same comparisons, as well as by utilizing information from the navigation system. If the DTED is being used for navigation by blending DTED and IMU data through a Kalman filter (as in the SITAN algorithm, for example), the filter may contain estimates of the DTED errors. These error estimates may be used to weight the DTED with respect to the sensed data. If an independent algorithm for LAM registration is implemented, the history of adjustments to the map position can provide a measure of confidence in the DTED. Measures of confidence are thus readily available for both sensor-derived and DTED terrain information and should be used.

The blending of sensed data and DTED must accommodate the relative resolutions available from the two sources. With the digital data quantized to 100 meters horizontally, a mismatch will occur for almost all sensors at some range. The association of sensed data with mapped data will thus vary from a radar return being associated with a part of a map cell, to being associated with several. Since the sensor resolution cell is, in general, not aligned with the LAM cells, overlap may occur because of the geometry of the situation.

Particular care must be given to the blending of data from small (but tall) features such as towers. Tower identification should be included in the digital data base so that the blending algorithm can recognize the presence of an obstacle which is unseen by the sensor, independent of the confidence estimated in the sensor data. Correspondingly, a detected, but unmapped, tower should be inserted in the LAM independent of the history of correctness of the DTED.

4.5 SENSOR CONTROL

The use of DTED allows an active sensor to be used more sparingly than if the sensor were the only source of data. The sensor's peak power may be reduced drastically, since ranges of less than five kilometers are required to detect unmapped obstacles in time for avoidance. This reduction in peak power corresponds to a consequent reduction in detection range by intercept receivers. Since the sensor data is blended with the DTED and maintained in memory, continuous operation of the sensor is not required. Intermittent, irregular operation of the sensor provides additional difficulty for an intercept system. Because the projected curved flight path is known, the sensor may be controlled to examine only the terrain of interest, rather than having to scan over the extent of all possible maneuvers of the aircraft.

The projection of the planned flight path onto the high resolution LAM immediately ahead of the aircraft allows determination of which cells have been examined by the sensor and which must be scanned to ensure flight safety. The availability of the DTED allows a-priori estimation of the sensor horizon, enabling a minimization of radiated energy into space. The DTED also prevents fruitless attempts to sense terrain obscured by nearby features, again reducing the radiated signature.

In addition to managing active sensors, passive sensor management may be accomplished. This can reduce aircraft detectability if the passive sensor can be "hidden," with reduced observability, when not in use.

4.6 FLIGHT TEST RESULTS (Reference 5)

A terrain following/terrain avoidance/threat avoidance system incorporating many of the features described above has been developed and demonstrated in flight. Texas Instruments, under contract to the United States Air Force Wright Aeronautical Laboratories, Avionics Laboratory, conducted an Enhanced Terrain Masked Penetration (ETMP) program from September 1983 through May 1986. This program demonstrated the feasibility of automated real time terrain following, terrain avoidance, and threat avoidance. Sixty flights were made in a modified Convair 580 aircraft, flying principally in an area where terrain following radars are habitually tested. The first equipment check-out flight was on 31 October, 1984 and the final formal demonstration flight occurred on 14 May 1986. The first flight under direction of the lateral path generation algorithm was in March 1985.

Typical demonstration flights involved a closed course with portions of the flight path both parallel and perpendicular to a series of ridges. Flights with and without simulated threats were made to illustrate the effectiveness of the system in using terrain masking to avoid exposure to hostile air defense systems.

Data collected on some flights, using a highly accurate Integrated Navigation System (an IMU tightly coupled to a Global Positioning System receiver and processor) were used to evaluate the accuracy of the DTED. Discrepancies on the order of 50 meters were not uncommon, with the DTED being both higher and lower than the measured terrain at different places in the 100 by 100 kilometer map area. One significant hill, some 300 meters above the surrounding terrain, was included in the map but did not exist in the real world. This experience has strengthened our belief that digital terrain data alone is not sufficient for safe low-level flight.

5. REFERENCES

The material presented herein has been drawn from a variety of sources and is believed to be representative of the consensus at researchers in this topic over the past decade; namely, that the advent of stored digital map data, on-board mission planning systems, and precise terrain-aided navigation systems can significantly improve the survivability of attack aircraft while penetrating hostile territory. Entry into the available literature can be gained through the following:

Cited References

1. Kupferer, R. A., and Halski, D. J., "Tactical Flight Management - Survivable Penetration", NAECON Proceedings, May 1984.
2. Murphy, W. J., "Simulation and Flight Test Results of Terrain Following/Terrain Avoidance and Terrain Correlation Navigation Systems Using Stored Terrain Data", AGARD Lecture Series No. 122, March, 1983.
3. Boozer, D. D., et al, "The AFTI/F16 Terrain-Aided Navigation System", Proceedings of the IEEE National Aerospace and Electronics Conference", May 1985, pp 351-357.
4. Dale, R. S., "TERPROM, Terrain Profile Matching", British Aerospace, Naval and Electronics System Division, July, 1987.
5. Chapoton, Charles W., Jr., "Artificial Intelligence in the Enhanced Terrain Masked Penetration Program", Defense Science and Electronics, November 1986, pp 62-72.

References of General Interest

1. GM Barney, "Enhanced Terrain Masked Penetration Program", NAECON Proceedings, May 1984.
2. Bird, M. W., "PNCS - A Commercial Flight Management Computer System", AIAA Guidance and Control Conference, August 1982.
3. Bird, M. W., et al, "The Trajectory Generator for Tactical Flight Management", NAECON Proceedings, May 1983.
4. Harrington, W. W., Gates, T. G., Russ, D. E., "TF/TA System Design Evaluation Using Pilot-in-the-Loop Simulation: The Cockpit Design Challenge", SAE Aerospace Congress and Exposition, October 1984.
5. Hostetler, L. D., "A Kalman Approach to Continuous Aiding of Inertial Navigation Systems Using Terrain Signatures", IEEE Milwaukee Symposium on Automatic Computer Control Proceedings, April 1976, pp 305-309.
6. Hostetler, L. D., "Nonlinear Kalman Filtering Techniques for Terrain-Aided Navigation", IEEE Transactions on Automatic Control, Vol. AC-28, March 1983.
7. Johnson, M. W., "Analytical Development and Test Results of Acquisition Probability for Terrain Correlation Devices Used in Navigation Systems", AIAA Tenth Aerospace Sciences Meeting, AIAA Paper No. 72-122, 1972.
8. Katt, D. R., Wendl, M. J., Young, G. D., "Integrated GPS, DLMS, and Radar Altimeter Measurements for Improved Terrain Determination", Institute of Navigation, US Air Force Academy, June 1982.
9. Maroon, G. N., McDonough, W. G., and Maschek, T. J., "Tactical Flight Management - Total Mission Capability", NAECON Proceedings, May 1984.
10. Murphy, W. J., "Integrated Flight/Weapon Control Concepts", NAECON Proceedings, May 1980.
11. Murphy, W. J., and Young, W. L., Jr., "Integrated Flight/Weapon Control Design and Evaluation", NAECON Proceedings, May 1981.
12. Murphy, W. J., and Young, W. L., Jr., "Tactical Flight Management", ADPA/AIAA Technical Meeting on Avionics Technology and Systems Development, Nellis AFB, NV, December 1982.
13. Murphy, W. J., and Young, W. L., Jr., "Tactical Flight Management - System Definition", IEEE Position, Location, and Navigation Symposium, Atlantic City, NJ, December 1982.
14. Wall, J. E., Jr., Wendl, M. J., and Young, G. D., Jr., "Advanced Automatic Terrain Following/Terrain Avoidance Control Concepts - Algorithm Development", AIAA Guidance and Control Conference, Albuquerque, NM, August 1981.
15. Wendl, M. J., Katt, D. R., "Advanced Automatic Terrain Following/Terrain Avoidance Control Concepts Study", NAECON Proceedings, May 1982.

ACKNOWLEDGEMENTS

The authors wish to acknowledge the significant contributions of our colleagues at McDonnell Aircraft, Texas Instruments and elsewhere. The brief survey presented herein barely touches upon the progress made in this field during the 1980's by the entire aerospace community.

PART VI

Civil Aircraft Navigation and Traffic Control

INDEPENDENT GROUND MONITOR COVERAGE OF GLOBAL POSITIONING SYSTEM (GPS) SATELLITES FOR USE BY CIVIL AVIATION¹

by

Karen J. Viets, Member of the Technical Staff
The MITRE Corporation, MS W-295
7525 Colshire Drive
McLean, VA 22102-3481
United States

ABSTRACT

The Federal Aviation Administration plans to independently monitor signals-in-space from the Global Positioning System (GPS) for the purpose of providing immediate awareness to civil aviation users of the operational status of GPS when it is used in the National Airspace System. The operational status will be disseminated to Air Traffic Control and will possibly be broadcast from ground monitoring stations to GPS aviation users via a dedicated integrity channel. This report describes an algorithm that measures the coverage of a configuration of ground monitoring station locations, and applies the algorithm to several different configurations of ground monitoring stations to compare the coverage provided. Also included are the resulting ground monitoring station configurations that provide the best coverage of GPS signals for several specific geographical areas, the conterminous United States (CONUS), Canada, and Alaska.

INTRODUCTION

The Global Positioning System (GPS) is being considered by the Federal Aviation Administration (FAA) for use in nonprecision approach guidance in the National Airspace System (NAS). The FAA has determined, however, that a method for assuring the integrity of the GPS signals-in-space must be developed to meet the 10 second pilot notification time currently required for nonprecision approach navigation aids. One method for providing the necessary GPS integrity is to use an independent ground monitoring network along with a channel to broadcast the GPS signal status to the aviation users [1].

Such a ground monitoring network includes a number of ground monitoring stations that receive the signals from the GPS satellites in view. The network also includes several integrity control nodes, collocated with some of the ground monitoring stations, which process GPS signals-in-space to determine their status. GPS signal status messages are prepared at the integrity control nodes and then broadcast to the GPS users in the NAS, most probably via a geostationary satellite based integrity channel.

An important element in assessing the feasibility of implementing such a network is to determine how many ground monitoring stations to use, and where these stations should be placed in order to minimize the number of stations necessary to provide the required coverage of the geographical area(s) of interest. This paper develops an algorithm for measuring the coverage of a configuration of ground monitoring station locations, and for comparing different configurations of ground monitoring stations to determine which configuration provides the best GPS coverage for a given area of the world. The algorithm is applied to the following geographical areas of GPS use: the conterminous United States (CONUS), Canada, and Alaska.

DEVELOPMENT OF CANDIDATE GROUND MONITORING STATION CONFIGURATIONS

The ground monitoring station configuration that provides the required coverage of a given area with the smallest number of stations satisfying the operational and institutional constraints is the best configuration for the GPS ground monitoring network. This section describes the requirements for the ground monitoring station locations and identifies the locations that are considered in this analysis.

User Versus Monitor Visibility of GPS Satellites

One method for comparing the extent of GPS satellite visibility for a given area of use to the GPS satellite visibility for a given configuration of ground monitoring stations is to first determine the area at the altitude of the GPS orbit that can be "seen" from a specific location on the earth. Figure 1 shows this area in the GPS orbit and illustrates how it can be projected back down to the earth, forming a cone with the earth's center. Assuming that the earth is spherical, the area that is seen at the altitude of the GPS orbit (10,898 nautical miles) is directly proportional to the area that has been projected down to the earth's surface. Therefore, the area seen at GPS altitude may be represented by the area projected onto the earth by a satellite that is located 10,898 nautical miles above location L on the surface of the earth [2], [3]. The computations for determining the area on the surface of the earth that is covered by a satellite in GPS orbit are described in detail in the Appendix.

Given an area of GPS use such as CONUS, it is necessary to determine where to place a minimum number of ground monitoring stations to provide the required coverage of the GPS signals-in-space. If the area of GPS orbit that can be seen from a particular location on earth can be represented by a circle of visibility projected onto the earth, then a similar circle can be drawn for each prospective ground monitoring station as well as for each potential location of GPS use.

¹ This paper is based upon navigation system studies performed by MITRE for the Systems Engineering Service, Federal Aviation Administration under Contract No. DTF401-84-C-00001. The data presented herein do not necessarily reflect the official views or policy of the FAA.

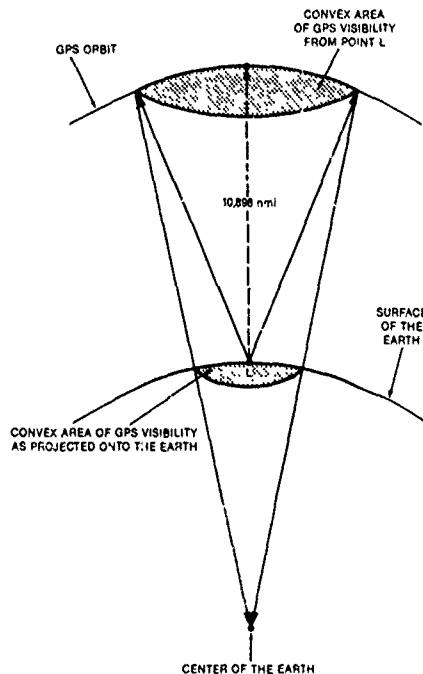


FIGURE 1
GPS VISIBILITY FROM A POINT ON THE EARTH

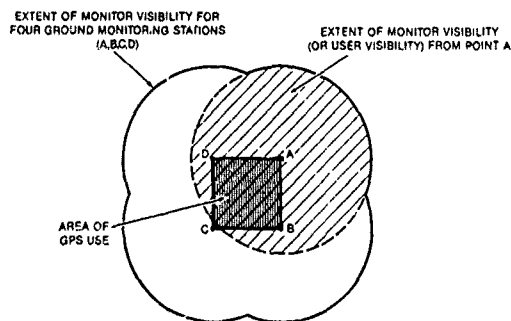


FIGURE 2
PLACEMENT OF GROUND MONITORING STATIONS FOR
COVERAGE OF A GIVEN AREA OF GPS USE

Placing the ground monitoring stations on the outer edges of the area of GPS use, at points that are furthest from the center of the area (e.g., corners of the square in Figure 2), allows the circles of monitor visibility to extend at least as far as the circles of user visibility.² For areas of GPS use like those considered in this analysis, the center of the area is covered by at least one ground monitoring station because the radius of GPS visibility is much greater than the distance from a monitoring station to the center of the area. Monitoring stations placed on the edges of a given area will therefore maximize the coverage for users within that area.

It is desirable to minimize the number of ground monitoring stations for a given area of use. To address this, a number of different candidate configurations of ground monitoring stations are tested to see how well they cover each area of GPS use. The ground monitoring station configuration that provides the required coverage with the smallest number of stations is chosen; however, it is possible that an even smaller number of stations will provide coverage almost as complete as that of the chosen configuration. In this case, a decision must be made as to whether or not the additional coverage of the chosen configuration is worth the cost of implementing and maintaining more stations.

² Recent analysis shows that the ground monitoring antennas are indeed capable of receiving GPS signals at minimum elevation angles that are equal to or less than those achieved by the user antennas.

Operational and Institutional Constraints

Also to be considered are the operational and institutional constraints involved in establishing and operating a ground monitoring station. Two of these constraints are described below.

- Air Route Traffic Control Centers (ARTCCs) serve as an excellent ground monitoring station location for the reason of access to a data communications network (e.g., NADIN), as well as for the reasons of operation and maintenance.
- The ground monitoring stations may possibly require the GPS Precision Code. If this is the case, it is essential that the facilities chosen for ground monitoring stations be secure.

Considering the above constraints, it is desirable to locate the ground monitoring stations at major air traffic control facilities, such as the ARTCCs in CONUS. This consideration, along with the performance requirements concerning the placement and number of ground monitoring stations, will be used in determining the candidate ground station configurations to be studied in this analysis.

Candidate Ground Monitoring Station Configurations

The areas of GPS use analyzed in this paper include CONUS, Canada, and Alaska. Several different configurations of candidate ground monitoring stations have been chosen near the edges of these areas using the *Atlas of the World* [4] to determine a feasible configuration that will provide the most complete coverage for each area. The algorithm developed in this paper is applied to each of the candidate configurations to produce quantitative results. Both the performance requirements and the operational and institutional constraints were considered in choosing the candidate configurations of ground monitoring stations.

The ground monitoring station configurations that are analyzed in this paper include the following:

- Monitor stations located in CONUS only: Miami, Los Angeles, Seattle, and Boston
- Monitor stations located in CONUS and Canada: Miami, Los Angeles, Vancouver, Frobisher Bay, and Halifax [5]
- Monitor stations located in CONUS, Canada, and Alaska: Miami, Los Angeles, Barrow,³ Frobisher Bay, and Halifax
- Monitor stations located in CONUS, Canada, Alaska, and Hawaii:⁴ Miami, Los Angeles, Honolulu, Anchorage, Frobisher Bay, and Halifax

A map of the western hemisphere that shows each of the candidate locations of ground monitoring stations is shown in Figure 3.

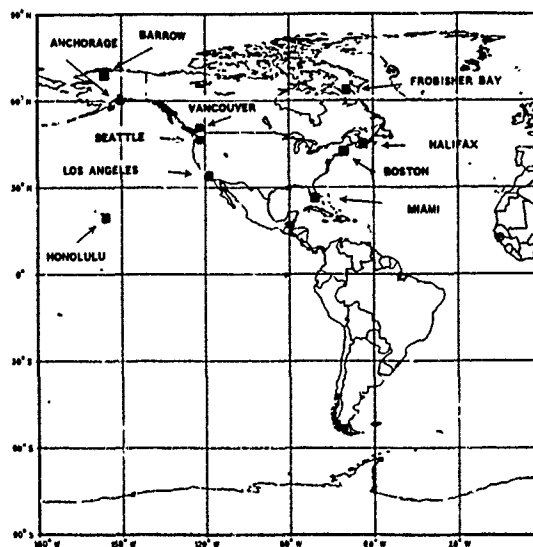


FIGURE 3
CANDIDATE LOCATIONS FOR GROUND MONITORING STATIONS

³ Although Barrow, Alaska, is not a major air traffic control facility, it was included in this study to illustrate the benefit of a more northern ground monitoring station.

⁴ The station in Hawaii has been chosen to illustrate the benefit of a ground monitoring station outside the boundary of CONUS. Ground monitor coverage is increased because California extends further west than either Seattle or Los Angeles.

ANALYSIS OF DIFFERENT GROUND MONITORING STATION CONFIGURATIONS FOR AN AREA OF GPS USE

This section describes the algorithm developed to assess the monitoring coverage of proposed ground monitoring station configurations. First, the individual areas representing the area of GPS orbit visible to each user and each ground monitoring station are calculated as illustrated in Figure 1. The individual areas of GPS visibility for the users in a particular geographical area are combined to determine the extent of visibility for the users in that area. In a similar manner, the areas visible to each ground monitoring station are combined to give the extent of visibility for that particular ground monitoring station configuration. Comparing the extent of user visibility with the extent of monitor visibility, monitoring holes that exist in the coverage are places where the users can view GPS satellites when the ground monitoring stations cannot view them. A satellite may not be used for certain phases of navigation (e.g., nonprecision approach) when it passes through a monitoring hole because its integrity status message will not be transmitted.

Because a monitoring hole itself is not significant unless a GPS satellite passes through the hole, it is necessary to determine the amount of time that each satellite remains visible to the users while it is not visible to the ground monitoring stations. The total time that GPS satellites are visible to users and are not visible to at least one ground monitoring station then becomes a quantitative measure of comparison for the different configurations of stations. Statistical values of these total times are compared.

The Appendix describes the technical details of this analysis. The analysis results for the three previously specified areas of GPS use are described in the next section.

RESULTS FOR SPECIFIC AREAS OF GPS USE

The analysis of the monitoring coverage provided by a given configuration of ground monitoring stations, described in the Appendix, was applied to CONUS, Canada, and Alaska. For all of the computations made in this analysis, the GPS constellation consists of the planned 18 satellites in 3 orbits plus 3 operating spare satellites, and the minimum elevation angle achieved by the civil aviation user is assumed to be equal to that achieved by the ground monitoring station.

As mentioned in the previous section, the absence of an integrity signal disseminated for GPS satellites when they pass through monitoring holes prevents their use for certain phases of navigation. Only the users located outside the boundaries of the ground monitoring stations themselves are actually affected because they are the only users that can view the satellites when the monitoring stations cannot view them. For example, each of the ground monitoring stations in CONUS is located within the boundary of CONUS; therefore, users on the boundary of CONUS can view GPS satellites when the ground monitoring stations cannot view them. As a result, the users on the outer edges of CONUS benefit from ground monitoring station configurations that include stations outside the boundary of CONUS.

Although the monitoring holes only affect the users on the edges of the particular geographical area, the total time that a satellite spends in monitoring holes for a given configuration of ground monitoring stations does not apply continuously to all of the users in the edge of that area. For example, if the mean time that the GPS satellites are in monitoring holes is one hour, the average satellite will only be visible to a particular user for a small portion of that hour. The effect of the mean total time that each satellite spends in a monitoring hole is therefore distributed among the users on the outer edges of the geographical area.

Table 1 shows the results for each of the areas of GPS use with each candidate configuration of ground monitoring stations. In this table, N represents the number of GPS satellites that pass through monitoring holes. For these GPS satellites, the conditional mean, M, and standard deviation, σ , are given to describe the total amount of time, in hours, that these satellites are visible to the users while they are not visible to at least one ground monitoring station. Only the satellites that pass through monitoring holes are included in the mean, M, and the standard deviation, σ .

As seen in Table 1 for the GPS users in CONUS, all 21 GPS satellites pass through the monitoring holes for the configuration of ground monitoring stations located only in CONUS. The mean time that these 21 satellites spend in monitoring holes is on the order of an hour for this example. The shaded area on the map in Figure 4 represents the monitoring holes for this example, or the area where users in CONUS are able to view GPS satellites when they are not monitored by the stations in CONUS. These monitoring holes are distributed in time and location around the area near the outer edge of GPS user visibility. The mean M, although much larger than the actual time a user would be exposed to an unmonitored satellite, is assumed to provide a relative measure of the impact on users for comparison of different monitor configurations. Similar results for the other configurations of ground monitoring stations and areas of GPS use are also given in the table.

It can be seen from the table that the configurations including stations located in CONUS, Canada, and Alaska, as well as CONUS, Canada, Alaska, and Hawaii, limit the mean of the total amount of time that GPS satellites are not monitored to within 0.5 hour for the users on the perimeters of each of the individual areas. Because this total amount of time is distributed along the perimeters, either of these two configurations of ground monitoring stations provide almost continuous monitoring coverage of the GPS signals-in-space.

It can also be seen from the table that the ground monitoring station configuration that provides the most complete coverage for the GPS users in CONUS includes stations located in CONUS, Canada, Alaska, and Hawaii. For the users in Canada and Alaska, however, the mean total time that GPS satellites spend in monitoring holes is lowest for the configuration of ground monitoring stations located in CONUS, Canada, and Alaska. This result shows that Alaska and the northwestern part of Canada benefit from a more northern ground monitoring station such as Barrow, Alaska. In this case, it is necessary to consider the operation and maintenance of a station in Barrow as opposed to one in Anchorage.

TABLE 1
TOTAL TIME STATISTICS FOR GPS SATELLITES
THAT PASS THROUGH MONITORING HOLES
(APPLY ONLY TO PERIMETERS OF GEOGRAPHICAL AREAS)

CANDIDATE CONFIGURATION	MONITORING STATION LOCATIONS	REGION OF GPS USE		
		CONUS	CANADA	ALASKA
CONUS	Miami, Los Angeles, Seattle, Boston	N = 21 M = 0.92 hrs σ = 0.47 hrs	N = 21 M = 1.66 hrs σ = 0.78 hrs	N = 21 M = 0.89 hrs σ = 0.37 hrs
CONUS AND CANADA	Miami, Los Angeles, Vancouver, Frobisher Bay, Halifax	12 0.40 0.14	21 1.01 0.36	21 0.57 0.16
CONUS, CANADA, AND ALASKA	Miami, Los Angeles, Barrow, Frobisher Bay, Halifax	12 0.34 0.10	17 0.30 0.15	21 0.46 0.23
CONUS, CANADA, ALASKA, AND HAWAII	Miami, Los Angeles, Honolulu, Anchorage, Frobisher Bay, Halifax	8 0.25 = 0	17 0.43 0.17	19 0.49 0.22

N = Number of satellites that pass through monitoring holes
M = Mean time in hours that the N satellites spend in monitoring holes
 σ = Standard deviation in hours for the time that the N satellites spend in monitoring holes

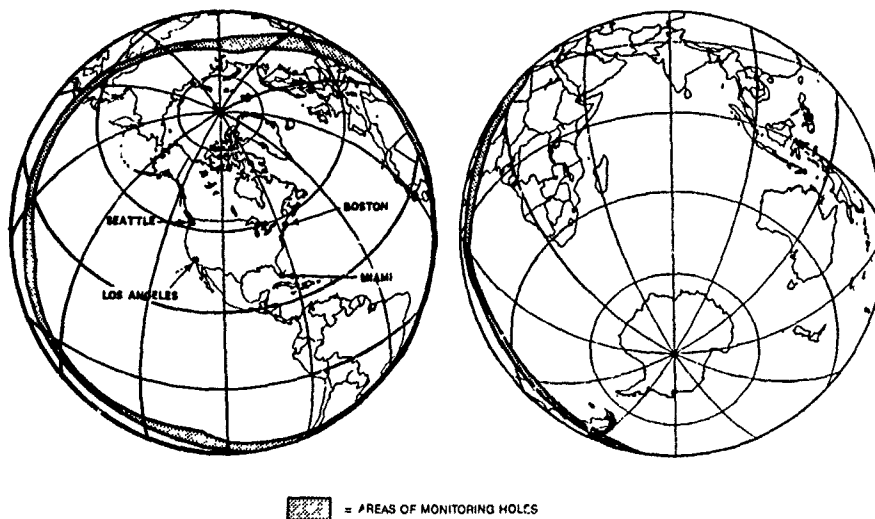


FIGURE 4
MONITORING HOLES FOR USERS IN CONUS,
GROUND MONITORING STATIONS IN CONUS

CONCLUSIONS

The algorithm developed and applied in this paper compares the coverage provided by candidate ground monitoring station configurations for specific areas of GPS use. This algorithm, based on the conservative assumption that the minimum elevation angle achieved by the civil aviation user is equal to that achieved by the GPS ground monitoring station, may be used as a tool in determining the optimum configuration of ground monitoring stations.

The results given here for the specific areas of GPS use, CONUS, Canada, and Alaska, show that the ground monitoring stations placed on the outer edges of the defined area of GPS use provide the most complete coverage of the GPS satellites. The configuration with ground monitoring stations located in CONUS, Canada, Alaska, and Hawaii provides the most complete coverage for GPS users in CONUS. Placing a ground monitoring station at Barrow, rather than in Anchorage, Alaska, enhances the monitoring coverage for GPS users in Canada and Alaska.

Recent analysis of the GPS Operational Control Segment (OCS) monitoring antennas indicates that the antennas are capable of receiving GPS signals at elevation angles that are equal to or less than those expected to be achieved by the aviation user antennas. A potential solution for eliminating the monitoring holes is for the ground monitoring stations to operate at a lower elevation angle than the civil aviation users. This would allow the visibility of the ground monitoring stations to extend further, possibly providing full coverage for those civil aviation users on the outer edges of the geographical area of GPS use.

APPENDIX COMPUTER SOFTWARE DEVELOPED TO ANALYZE DIFFERENT CONFIGURATIONS OF GROUND MONITORING STATIONS

The algorithm described in the third section was applied to the specific areas of GPS use and the ground station configurations described in the second section. For each area of GPS use, the following steps were completed to determine the total amount of time that the GPS satellites spend in monitoring holes for each configuration of ground monitoring stations:

1. Determine the extent of visibility for the GPS users in the given geographical area
2. Determine the extent of visibility for the particular ground monitoring station configuration
3. Compare the edge of user visibility with the edge of monitor visibility to find the monitoring holes in the GPS coverage for that particular area of GPS use
4. Determine how much time, in hours, that each of the 21 GPS satellites spends in the monitoring holes for any 24 hour period (one complete ground track)
5. Evaluate the configuration of ground monitoring stations using statistical values and histograms

These steps were completed for each area of GPS use and each configuration of candidate ground monitoring stations. To aid in applying the algorithm, computer software was developed to make the necessary computations and to project the resulting areas of visibility onto a world map.

A computer program was developed to determine the extent of visibility for GPS use in a given geographical area, as well as to determine the extent of visibility for the candidate ground monitoring station configurations (Steps 1 and 2). The program computes the edges of the two resulting areas of visibility in geodetic (latitude and longitude) coordinates, given specific visibility points on the earth in the same coordinate system (see second section). The visibility points used for Step 1 are points that have been selected from the boundaries of the geographical area of use.^{*} The visibility points used in performing Step 2 are the locations of candidate ground monitoring stations.

This program computes the location of the area on the surface of the earth that represents the area of the GPS orbit seen from a particular location on the earth. This area is projected as a circle on the surface of the earth. The program also determines the geodetic coordinates that define the edge of the surface area. Shown below are the equations for computing the area on the surface of the earth; it is equivalent to the area covered by the satellite. Figure 5 illustrates the trigonometry involved. [2]

$$A_{cov} = (2\pi r_e^2)(1 - \cos \phi_e) \quad (1)$$

where

$$\begin{aligned} A_{cov} &= \text{area covered by satellite} \\ \phi_e &= \cos^{-1} [(r_e \cos \phi_2) / (r_e + h)] - \phi_2 \\ r_e &= 3,444 \text{ nmi} \\ h &= 10,898 \text{ nmi} \\ \phi_2 &= 7.5^\circ \end{aligned} \quad (2)$$

then

$$r_{cov} = r_e \phi_e \quad (\text{in nautical miles}) \quad (3)$$

To locate the edges of the individual circles of visibility in geodetic coordinates, it was necessary to compute points around the circles in angle increments, computing each edge coordinate pair given the following:

- The center coordinate pair for the circle of visibility (the visibility point)
- The geodesic radius of the circle of visibility
- The angle of the radial line relative to the North Pole (true bearing)

Spherical trigonometry and the use of the terrestrial triangle are utilized in the algorithm to determine the geodetic coordinates of each representative edge point, as shown in detail below; Figure 6 illustrates the geometry of the solution. [3]

^{*} When selecting the edge points for Canada, possible GPS use was considered up to 70 degrees north latitude.

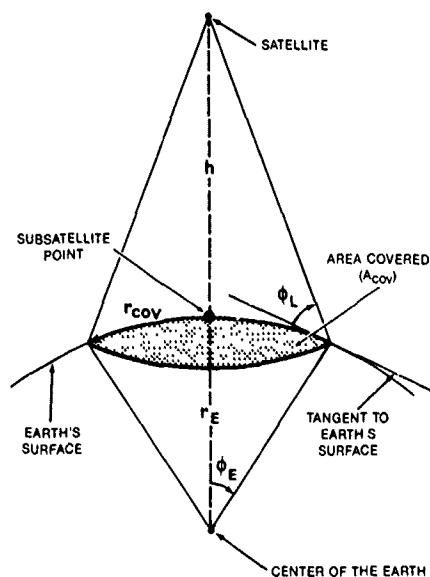


FIGURE 5
COMPUTING THE GPS VISIBILITY FROM A
LOCATION ON THE EARTH

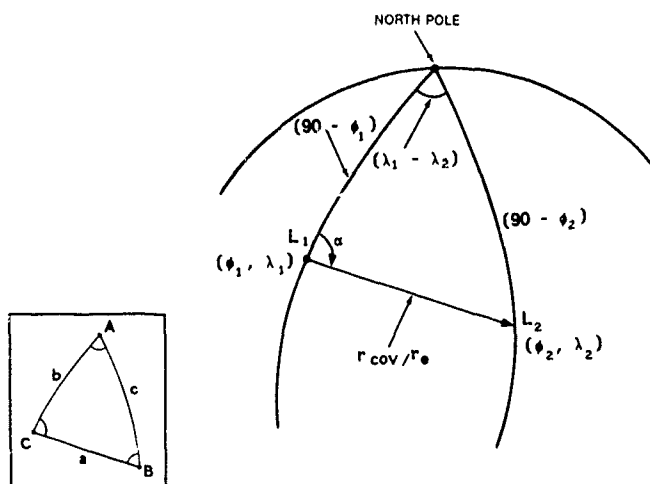


FIGURE 6
TERRESTRIAL TRIANGLE TO COMPUTE THE EDGE COORDINATES

Given: $a = r_{cov} / r_E$
 $b = (90 - \phi_1)$
 $C = \alpha$ (angles incremented around circle)

Find: $L_2 (\phi_2, \lambda_2)$

where ϕ = latitude and λ = longitude

Solve for the right hand side of the following equations

$$\frac{1}{2}(A - B) = \tan^{-1} [\cot \frac{1}{2}C \times \sin \frac{1}{2}(a - b) / \sin \frac{1}{2}(a + b)] \quad (4)$$

$$\frac{1}{2}(A + B) = \tan^{-1} [\cot \frac{1}{2}C \times \cos \frac{1}{2}(a - b) / \cos \frac{1}{2}(a + b)] \quad (5)$$

The sum of Equations (4) and (5) gives the value of the angle A; Equation (5) minus Equation (4) gives the value for angle B. Given, from Figure 6, that

$$\begin{aligned}\lambda_1 - \lambda_2 &= A, \text{ then} \\ \lambda_2 &= \lambda_1 - A\end{aligned}\quad (6)$$

The next step is to determine the leg length c as follows:

$$c = 2 \times \tan^{-1} [\tan \frac{1}{2}(a - b) \times \sin \frac{1}{2}(A + B) / \sin \frac{1}{2}(A - B)] \quad (7)$$

Also given from Figure 6 that

$$\begin{aligned}c &= 90^\circ - \phi_2, \text{ then} \\ \phi_2 &= 90^\circ - c\end{aligned}\quad (8)$$

This algorithm is repeated at constant intervals around the circle of visibility to determine the geodetic coordinates of the representative edge points.⁶ The program developed to perform the above calculations for each area of visibility also computes the geodetic coordinates of the subsatellite points that form the ground trace for each satellite. Plots of these points at time intervals of one half hour are used in Step 4 to determine how much time each satellite spends in the monitoring holes found in Step 3.

The plots of the extent of user visibility, the extent of monitor visibility (or monitor coverage), and the satellite ground traces are Lambert equal area projections produced by several Display Integrated Software System and Plotting Language (DISSPLA) programs. These programs use DISSPLA's plotting subroutines to input the geodetic coordinates for each of the above maps and plot them on a Lambert equal area projection of the world. Separate DISSPLA programs project the world coastlines and political boundaries on a Lambert equal area map of the world; the maps of user visibility, monitor coverage (or monitor visibility), and ground traces are used as overlays for the maps of the world.

REFERENCES

1. Braff, Ronald and Curtis Shively, "GPS Integrity Channel," *The Journal of the Institute of Navigation*, Vol. 32, No. 4, Winter, 1985-86, pp. 334-350.
2. Gagliardi, R. M., *Satellite Communications*, Belmont: Lifetime Learning Publications, 1984, p. 15.
3. Nielsen, K. L., *Modern Trigonometry, An Analytic Approach to Plane and Spherical Trigonometry*, New York: Barnes & Noble Books, 1966, pp. 202-204.
4. *Atlas of the World, Comprehensive Edition*, New York: Times Books, 1983.
5. Taylor, G. J., Private communication, Transport, Canada, March 1986.

ACKNOWLEDGEMENTS

The work reported in this paper was performed under the sponsorship of the Federal Aviation Administration's Systems Engineering Service, under the direction of Mr. Jerry W. Bradley, AES-310

⁶ Equations (4), (5), and (7) are equivalent forms of three of the four Analogies of Napier.

ANALYSIS OF THE INTEGRITY OF THE MICROWAVE LANDING SYSTEM (MLS) DATA FUNCTIONS

by

Dr M.B.El-Arini and M.J.Zeltser
The MITRE Corporation, Metrek Division
7525 Colshire Drive
McLean, Virginia 22102-3481
United States

ABSTRACT

The Microwave Landing System (MLS) transmits angle, data, and range information for use by airborne receivers. In this paper, the integrity of the data functions is analyzed in terms of the probability of undetected errors remaining in the data. The data format and integrity requirements were derived from the MLS standards and guidance material defined by the International Civil Aviation Organization (ICAO). Results show that the performance requirements can be met by: 1) averaging the received data bits of several samples of the same word using a majority voting; 2) reducing the bit error rate at the output of the receiver's decoder; and 3) a combination of the above techniques.

INTRODUCTION

The Microwave Landing System (MLS) ground-based equipments radiate signals from which the airborne receiver can determine azimuth and elevation angles, as well as digital data describing the ground system characteristics. Range information is also available using a slightly modified Distance Measuring Equipment (DME) technique which provides more accuracy than conventional DME.

Objective of the Paper

The objective of this paper is to present results of analyses performed on the integrity of the MLS data functions for use by MLS receiver manufacturers and standards development. The analysis includes a performance assessment and an identification of MLS receiver techniques to meet the integrity performance criteria. The performance of the data content in the MLS data functions is analyzed as well as the performance of the preamble (i.e., Barker code and the function identification code). The techniques addressed in this paper to meet ICAO-defined integrity performance requirements include: 1) averaging the received data bits of several samples of the same word using majority voting; 2) reducing the bit error rate at the output of the receiver's decoder; and 3) a combination of the above techniques.

Performance Requirements

The key measure of the validity of the data transmitted through the MLS data functions is the probability of undetected error (P_{ue}) remaining in the data after the decoding process in the MLS receiver. The requirements for P_{ue} as considered by the All Weather Operations Panel (AWOP) of ICAO [1, 2] will be used in this paper: namely, P_{ue} in acquired data (i.e., when receiver warnings are removed) should not exceed 10^{-6} at minimum signal power density and should not exceed 10^{-9} at critical points along the approach path.

Approach and Scope

This paper analyzes the integrity of the MLS data functions at the output of the MLS receiver. The analytic approach used in this paper is as follows:

1. Determine performance at the MLS coverage extremes.
2. Compare the performance from step 1 to the integrity requirements.
3. If the performance does not satisfy the integrity requirements, enhancement techniques will be identified.
4. Determine performance of enhancement techniques (if it is needed) which include:
 - a. Improve receiver sensitivity; and
 - b. Incorporation of signal processing in the airborne receiver.

Background

At a given MLS ground facility, angle and digital data transmissions are made on a single C-band frequency through the use of time division multiplexing. Just prior to each angle or data transmission, a unique digital function identification code is radiated to inform the receiver of the type of information to follow. In this manner, the transmissions of angle and digital data can be arranged in any time order without affecting receiver capability to decode the information. The range information is provided by Precision Distance Measuring Equipment (DME/P) using an interrogate/reply format which is compatible with conventional DME, and is performed independently of the angle and digital data in the L-band frequency.

This paper is based upon navigation system engineering studies performed by MITRE for the Federal Aviation Administration's (FAA's) Program Engineering and Maintenance Service. The contents of this paper reflect the views of the authors who are responsible for the facts and the accuracy of the data presented herein, and do not necessarily reflect the official views or policy of the FAA.

The MLS signal format for angle and data transmission is shown in Figure 1 [3,4]. Because each function is an independent entity in the time-multiplexed format, the receiver can accommodate any sequence of functions through use of a function preamble. This preamble is radiated in a specified coverage volume and contains an unmodulated radio frequency carrier acquisition period, a receiver synchronization (Barker) code, and the function identification (FID) code. The modulation of the digital portions of function transmissions is accomplished through differential phase shift keying (DPSK). Detection of the Barker code and the function identification code sets up the receiver circuitry to properly process the remainder of the function transmission. In the case of angle functions, the remainder of the transmission is comprised of successive scans of a highly directive, fan-shaped beam from which the receiver can measure the azimuth or elevation angle. For data functions, the rest of the transmission is composed of a stream of digital data modulated by the same DPSK techniques. Upon completion of the processing of a function transmission, the receiver awaits reception of the next function preamble whereupon the process is repeated.

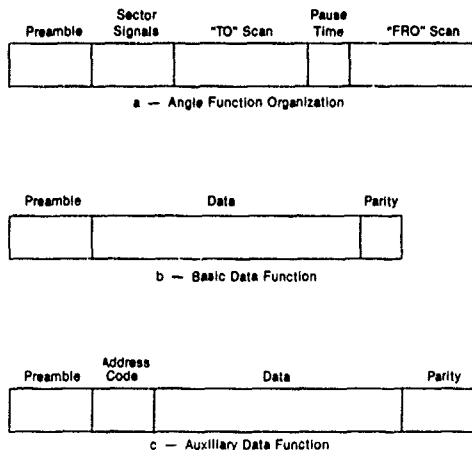


FIGURE 1
SIGNAL FORMAT FOR THE ANGLE FUNCTIONS AND
THE BASIC AND AUXILIARY DATA FUNCTIONS

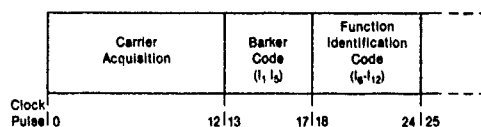
The MLS data is formatted in two word lengths and are called Basic and Auxiliary Data. The Basic and Auxiliary Data Functions have been incorporated into the MLS signal format to provide a means for describing the characteristics of a given MLS facility or approach procedure. Because the correctness of some of the data items to be provided on these functions can be critical to the safety of flight, the validity of the data is very important. Transmissions of basic data, such as the minimum glide path, are preceded by unique function identification codes which indicate what basic data items will follow. Auxiliary data are transmitted using one of three function identification codes [3] and an identifying address preceding the data portion of the word itself.

PERFORMANCE OF THE PREAMBLE

Every MLS function transmission begins with a preamble, as mentioned before, consisting of an unmodulated radio frequency carrier acquisition period, a Barker code, and an FID code as shown in Figure 2 [3 and 4]. Even though this digital preamble is used by the receiver only to control the processing of the rest of the function, preambles which are incorrectly decoded will have an impact on the processing of angle and data functions. In this section the performance of the preamble is examined in terms of the probability of detection and the probability of undetected error, as well as the effect of undetected errors in the preamble.

The initial part of the preamble (unmodulated carrier) is used to provide the reference or zero phase state of the binary data stream. The minimum duration required by the receiver is at least one bit length, but it may take longer depending on the ability of the phase lock loop (PLL) demodulator in the receiver to derive the reference state. The standard bit length in MLS is equivalent to the duration of one clock pulse (i.e., 64 μ s) and, as shown in Figure 2, 13 clock pulses (13 x 64 = 832 μ s) are provided for carrier acquisition. In the following analysis it is assumed that the unmodulated carrier is properly acquired (with the correct reference state) with very high probability even at the minimum signal-to-noise ratio (SNR) (i.e., the SNR that would be expected at the MLS coverage limit).*

*Experimental data has shown (Reference 1) that, at the minimum SNR, all words were decoded and no signal was lost due to the inability of the MLS receiver to derive the correct zero phase state.



Note. Duration of One Bit = Duration of One Clock Pulse = 64 μ s

FIGURE 2
PREAMBLE ORGANIZATION

Performance of the Barker Code

Barker codes are frequently provided in digital modulation to establish a timing reference for the data arriving at the receiver. In the MLS signal format, a 5-bit Barker code (11101) in bits I₁ to I₅ was chosen as a preferred compromise between probability of detection and false alarm considerations. The probability of detection (P_D) and the probability of error (P_{err}) of the five-bit Barker code are given by the following equations:

$$P_D = (1 - P_e)^n \quad (1)$$

$$P_{err} = 1 - P_D = 1 - (1 - P_e)^n \quad (2)$$

where $n = 5$ and P_e is the probability of an error in a single bit (sometimes called the bit error rate) and is given by [5] as

$$P_e = \frac{1}{2} [1 - \text{erf}(E/N_0)^{1/2}] \text{ for coherent PSK} \quad (3)$$

$$P_e = \frac{1}{2} \exp(-E/N_0) \text{ for DPSK} \quad (4)$$

where E is the average signal energy per received symbol, N_0 is the noise power density and $\text{erf}(\cdot)$ is the error function and is given by [5]:

$$\text{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x \exp(-t^2) dt$$

Note that P_D in Equation 1 and P_{err} in Equation 2 represent the fraction of the correct samples and the fraction of rejected samples at the output of the Barker decoder, respectively. P_D and P_{err} are computed for four values of P_e using Equations 1 and 2, respectively and are shown in Table 1. The $P_e = 2.7 \times 10^{-2}$ represents the bit error rate assumed for a "minimum" receiver at the 20 nmi coverage limit.* $P_e = 10^{-2}$, 10^{-3} , and 10^{-4} represent an approximate 1.3dB, 3.3dB, and 4.7dB improvement in the SNR** for DPSK detection from the minimum receiver case, respectively.

TABLE 1
PERFORMANCE OF THE BARKER DECODER

P_e	2.7×10^{-2}	10^{-2}	10^{-3}	10^{-4}
P_D	0.872	0.951	0.995	0.9995
P_{err}	0.128	0.049	0.005	0.0005

*The MLS power budget is based on a requirement that at least 1/4 percent of the transmitted preambles must be decoded at maximum range [3]. Using Equation 1 with $P_D = 0.72$ and $n = 12$ (which is the total length of the preamble), the corresponding bit error rate for the reception of the preamble is 2.7×10^{-2} .

**The improvement in the SNR is computed using the following equation for DPSK detector which is derived from Equation 4:
 $\Delta \text{SNR} = 10 \log_{10} [\ln(2P_e)/\ln(2P_e')] \text{ dB};$
 where P_e is the improved bit error rate and $P_e' = 2.7 \times 10^{-2}$ is the bit error rate for the minimum receiver at the MLS coverage limit.

Performance of the Function Identification Code

The FID consists of seven bits which follow the Barker code and indicates which MLS function is being radiated (see Figure 2). The FID is made up of five information bits (I_6 to I_{10}) allowing identification of 32 different functions, plus two parity bits (I_{11} and I_{12}). A total of 15 functions have already been established and identified [1, 3, and 4]. The parity bits were included to reduce the probability of improper processing of functions due to errors in the FID code. Parity Equations 5 and 6 represent the simplest form of parity that can detect single errors and double adjacent errors which are most likely to occur with DPSK modulation.

$$C_1 = I_6 + I_7 + I_8 + I_9 + I_{10} + I_{11} = 0 \quad (5)$$

$$C_2 = I_6 + I_8 + I_{10} + I_{12} = 0 \quad (6)$$

where "0" in the right hand side of Equations 5 and 6 means "even" and "+" means modulo 2 addition (i.e., $0 + 0 = 0$, $0 + 1 = 1$, $1 + 0 = 1$, $1 + 1 = 0$). It should be noted that Equation 5 detects all odd number of errors which occur in the first six bits of the FID code (i.e., bit numbers 6 to 11). In addition, Equation 6 detects some of the errors which are not detected by Equation 5 including double adjacent errors and, more generally, when the errors are manifested in one of the following ways: 1) the number of errors in the seven bits is odd, bit number 12 is in error, and the summation of the locations of the errors is even; or 2) the number of errors in the seven bits is even, bit number 12 is correct, and the summation of the locations of the errors is odd.

The probability of correctly decoding one FID code (P_D) and the probability of error (P_{err}) can be computed using Equations 1 and 2, respectively with $n = 7$. The probability of undetected error, P_{UE} , and the probability of detected error, P_{DE} , can be computed using Equations 7 and 8, respectively.

$$P_{UE} = \sum_{i=2}^n A_i P_e^i (1-P_e)^{n-i} \quad (7)$$

$$P_{DE} = P_{err} - P_{UE} \quad (8)$$

where

P_{UE} = The probability of undetected error in the FID code
 P_{DE} = The probability of detected error of one FID code
 A_i = The number of all possible combinations of i errors which are not detected by Equations 5 and/or 6
 P_e = The bit error rate as defined by Equations 3 and 4
 n = The length of each FID code
 = 7 bits
 i = The number of undetected errors

Note that the undetected error of a specific FID code may be decoded as another FID code, as discussed in the next section. To compute the probability of undetected error, Equation 7 is used where A_i is computed using the following reasoning:

1. When i is even (i.e., 2, 4, ...), the undetected errors using Equations 5 and 6 occur when:
 - a) there is no error in the second parity bit (i.e., bit no. 12); and b) the errors occur in even-even locations or odd-odd locations or in general the summation of all the locations of the i errors is even.
2. When i is odd (i.e., 3, 5, ...), the undetected errors using Equations 5 and 6 occur when:
 - a) there is an error in the second parity bit (i.e., bit no. 12); and b) the rest of the errors (i.e., $i - 1$ errors) occur in odd numbers of even locations and/or odd numbers of odd locations or in general the summation of all the locations of the $(i - 1)$ errors is odd.

The value of A_i as a function of the number of errors, i , is shown in Table 2. P_D , P_{err} , P_{UE} , and P_{DE} are computed for four values of P_e and are shown in Table 3. From this table it is seen that when $P_e = 2.7 \times 10^{-2}$ (i.e., for the minimum receiver at the MLS coverage limit) 83 percent of the samples of any specific FID code will be decoded correctly and 17 percent of the samples will be in error. Of those having errors, the FID decoder will reject about 98 percent using the parity Equations 5 and 6. The errors in the remaining 2 percent, however, will not be detected and those samples will be erroneously processed as a different FID.

TABLE 2
THE NUMBER OF COMBINATIONS OF i BIT ERRORS NOT DETECTED
BY THE FUNCTION IDENTIFICATION DECODER

NUMBER OF BITS HAVING ERRORS i	NUMBER OF COMBINATIONS NOT DETECTED BY PARITY A_i
2	6
3	9
4	9
5	6
6	0
7	1

TABLE 3
PERFORMANCE OF THE FUNCTION IDENTIFICATION CODE

P_e	2.7×10^{-2}	10^{-2}	10^{-3}	10^{-4}
P_D	0.8256	0.9321	0.9930	0.9993
P_{err}	0.1744	6.793×10^{-2}	6.979×10^{-3}	6.998×10^{-4}
P_{UE}	3.979×10^{-3}	5.763×10^{-4}	5.979×10^{-6}	5.998×10^{-8}
P_{DE}	0.1704	6.736×10^{-2}	6.973×10^{-3}	6.997×10^{-4}

The impact of a false FID decode is a function of the receiver processor design and the particular function being decoded and is presented in Reference 6. At worst, it causes the receiver to attempt to process the function improperly, but that should be rejected by the validation algorithms used for each function. Thus, the receiver might lose the ability to process valid functions for a period of time equal to the time duration of the falsely decoded function.

INTEGRITY OF THE BASIC DATA FUNCTION

The Basic Data Function (BDF) is intended to uplink the data required by the airborne system for basic angle processing and to support the straight-in approach. Examples are the MLS proportional coverage limits and the minimum glide path and other items as shown in References 3, 4, and 7. The organization of basic data words is defined by the ICAO [3] and the FAA [4]. The word length is 32 bits consisting of 12 bits for a preamble, 18 bits for information, and 2 bits for parity, as shown in Figure 3.

In this section the performance of potential validation techniques for data provided on the BDF is presented. Validation performance is quantified in terms of the probability of correctly decoding valid data, and the probability of undetected errors remaining in the data. Of the various ways performance could be improved over the baseline (only simple parity check), the techniques analyzed here include: 1) using additional airborne processing of multiple samples of a data word; 2) reducing the prevailing bit error probability; and 3) a combination of these two approaches.

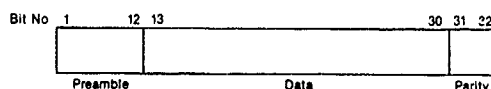


FIGURE 3
BASIC DATA WORD ORGANIZATION

Performance Using Simple Parity Checks

In this section, the baseline performance is quantified for various bit error probabilities in terms of the probability of correctly decoding valid data (probability of detection) and the probability of an undetected error. These performance measures are based on one sample of a basic data word where the simple parity equations are the only means used to detect errors in the single sample of that data word. The parity equations for the twenty bits of the basic data which follow the preamble, as specified in [3 and 4], are:

$$C_1 = I_{13} + I_{14} + \dots + I_{31} = 0 \quad (9)$$

$$C_2 = I_{14} + I_{16} + \dots + I_{32} = 0 \quad (10)$$

where "0" in the right hand side of Equations 9 and 10 means "even" and "+" means a modulo 2 addition.

The probability of correctly decoding valid data, P_D , of one basic data word can be computed using Equation 1 where $n=20$. The probability of error, P_{err} , in one word can be computed using Equation 2. P_{err} is again comprised of two parts: the probability of undetected error (P_{ug}), and the probability of detected error (P_{pg}). P_{ug} and P_{pg} can be computed using Equations 7 and 8, respectively where A_i is the number of all possible combinations of i errors which are not detected by Equations 9 and/or 10. Table 4 shows the values of A_i for $i=2$ to 5 which represents the dominant terms in the series of Equation 7 when P_e is relatively small. This table was generated by computer using similar reasoning as explained before.

TABLE 4
THE NUMBER OF COMBINATIONS OF 1 BIT ERRORS
NOT DETECTED BY BASIC DATA PARITY CHECKS

NUMBER OF BITS HAVING ERRORS i	NUMBER OF COMBINATIONS NOT DETECTED BY PARITY A_i
2	81
3	90
4	1956
5	1920

An upper bound of the probability of undetected error can be computed using the following formula [8]:

$$P_{UE} \leq \frac{1}{2} \sum_{i=2,4,\dots}^n \binom{n}{i} P_e^i (1-P_e)^{n-i} \\ = \frac{1}{2} \left\{ \frac{1}{2} [1 + (1-2P_e)^n] - (1-P_e)^n \right\} \quad (11)$$

This upper bound is based on the assumption that 50 percent of the errors not detected by the first parity equation (Equation 9), which represents all even number of errors (i.e., $i = 2, 4, 6, \dots$), are also not detected by the second parity equation (Equation 10).

P_D , P_{err} , P_{UE} and P_{DE} are computed for four values of P_e using Equations 1, 2, 7, 8, respectively with $n=20$, and are shown in Table 5. The $P_e = 2.7 \times 10^{-2}$ represents the bit error probability assumed for a "minimum" receiver at the 20 nmi coverage limit as discussed before. $P_e = 10^{-2}$, 10^{-3} and 10^{-4} represent an approximate 1.3dB, 3.3dB and 4.7dB improvement in the SNR for DPSK detector from the minimum receiver case, respectively. From Table 5 it is seen that the receiver processing of a single data sample will satisfy the $P_{UE} \leq 10^{-6}$ requirement only if the bit error rate $P_e \leq 10^{-4}$.

TABLE 5
BASELINE PERFORMANCE OF BASIC DATA WORD PARITY CHECKS

P_e	2.7×10^{-2}	10^{-2}	10^{-3}	10^{-4}
P_D	0.5784	0.8179	0.9802	0.998
P_{err}	0.4216	0.1821	1.981×10^{-2}	1.9981×10^{-3}
P_{UE}	3.789×10^{-2}	6.852×10^{-3}	7.964×10^{-5}	8.0863×10^{-7}
P_{DE}	0.3837	0.1752	1.973×10^{-2}	1.9973×10^{-3}

Performance Using Majority Voting in Addition to Simple Parity Checks

One airborne receiver processing method which could be used to reduce the probability of an undetected error is the majority voting. The block diagram of the majority voting followed by a decoder is shown in Figure 4. This detector uses a majority voting scheme, i.e., it checks N samples of the same bit of specific word and takes the majority state. Thus, at least m samples out of N should be the same where m is greater than $N/2$. Once the data are processed by the detector, a decoder which applies parity Equations 9 and 10 could be used. The main reason for using a majority voting is to improve the effective bit error rate of data going into the decoder without requiring SNR improvements. However, this comes at the expense of acquisition time as the receiver must receive multiple samples of the same word to complete validation.

The new bit error rate, P_{el} , at the output of the detector can be computed in terms of the prevailing bit error rate, P_e , of received data by:

$$P_{el} = \sum_{k=m}^N \binom{N}{k} P_e^k (1-P_e)^{N-k}, \text{ where } m > N/2 \quad (12)$$

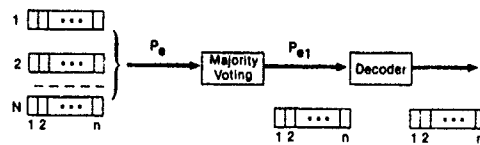


FIGURE 4
MAJORITY VOTING

Using Equation 12, P_{e1} is plotted against P_e and is shown in Figure 5 for the following cases of (m/N) : (2/3), (3/4), (3/5), and (4/5).

As evident in this figure, to achieve the desired effective bit error rate $P_{e1} \leq 10^{-4}$ (in order to satisfy the requirement for $P_{UG} \leq 10^{-6}$), the bit error rate of data going into the detector, P_e , should be less than or equal to the following values:

- o 5.8×10^{-3} for $m = 2, N = 3$
- o 3.0×10^{-2} for $m = 3, N = 4$
- o 2.2×10^{-2} for $m = 3, N = 5$
- o 7.0×10^{-2} for $m = 4, N = 5$

If 3-out-of-4 or 4-out-of-5 detectors are used, the requirement for $P_{UG} \leq 10^{-6}$ at the MLS coverage limit will be satisfied with the bit error rate assumed for the minimum receiver and, therefore, would not require improvements in the SNR. For the other two cases (i.e., 2-out-of-3 and 3-out-of-5 detectors) the requirement cannot be met without SNR improvements. If the requirement for the bit error rate at the MLS coverage limit for the minimum receiver was 2.7×10^{-2} , then the 3-out-of-4 and 4-out-of-5 detectors would be required.

In summary, any point shown in the shaded solution space of Figure 5 should satisfy both requirements (i.e., $P_{e1} \leq 10^{-4}$ and $P_e \leq 2.7 \times 10^{-2}$) at the MLS coverage limit. Any point on the $P_e = 2.7 \times 10^{-2}$ line that has $P_{e1} \leq 10^{-4}$ will satisfy the $P_{UG} \leq 10^{-6}$ requirements without SNR improvement (e.g., as in the 3-out-of-4 and 4-out-of-5 detectors cases). Otherwise any point within the solution space needs a SNR improvement to satisfy these requirements. These results are summarized in column 2 of Table 6.

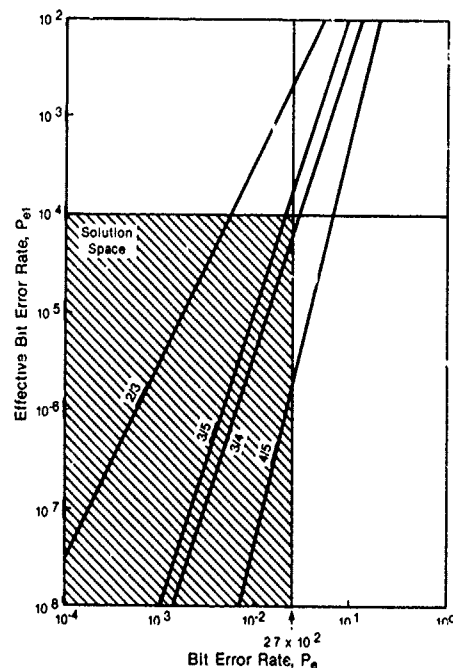


FIGURE 5
THE EFFECTIVE BIT ERROR RATE, P_{e1} , VERSUS THE
BIT ERROR RATE, P_e , FOR BASIC DATA WITH
MAJORITY VOTING

TABLE 6
EXAMPLE TECHNIQUES TO ATTAIN
 $P_{UE} = 10^{-6}$ FOR MLS BASIC
 DATA AT LIMITS OF COVERAGE⁽¹⁾

RECEIVER PROCESSING		APPROXIMATE BIT ERROR RATE TO ATTAIN $P_{UE} = 1 \times 10^{-6}$	APPROXIMATE SNR IMPROVEMENT NEEDED IN THE MIN. RECEIVER TO ACHIEVE THE BIT ERROR RATES IN COLUMN 2	TRANSMISSIONS REQUIRED TO ACQUIRE A BASIC DATA WORD (0.95 PROBABILITY) (see Reference 6)
Parity Check ⁽²⁾ (2-Parity Bits)		0.0001	4.7 dB	1
Majority Voting ⁽³⁾ Plus Parity Check	m/N			
	2/3	0.0058	1.8 dB	4
	3/4	0.0270	0.0 dB	8
	3/5	0.0220	0.3 dB	9

- Notes: 1) All values are approximate.
 2) Provided in existing basic data word format.
 3) Majority voting performs m out of N voting on each bit prior to the parity check.

It should be noted that when the requirement for $P_{UE} \leq 10^{-6}$ is satisfied at the extremes of coverage, the other requirement for $P_{UE} \leq 10^{-9}$ at critical regions on the approach path should be easily achieved primarily due to improved SNR. If the critical region is defined as the area within 10 nmi of the runway (as proposed in Reference 1), the SNR improvement over the case at the MLS coverage limit is at least 6dB. With this improvement, the bit error rate becomes very small ($P_e \leq 4.493 \times 10^{-6}$) and results in $P_{UE} \leq 1.64 \times 10^{-9}$ based on one sample using Equation 7 with A_i as given in Table 4. A majority voting would make P_{UE} even smaller than 10^{-9} in the critical regions.

The number of transmissions required to acquire a basic data word, when the majority voting is used, is computed using Equation 13. (See Reference 6).

$$P_{acq} = \text{Prob} (k \geq M)$$

$$= \sum_{k=M}^{NT} \binom{NT}{k} P_D^k (1-P_D)^{NT-k} \quad (13)$$

where

P_{acq} = The probability of the airborne receiver acquiring the necessary number of data word samples (in this paper, the minimum acceptable P_{acq} is assumed to be 0.95),

P_D = The probability of detecting one preamble,

M = The number of samples which must be received to satisfactorily reduce the probability of an undetected error (i.e., through majority voting),

NT = The number of transmitted samples.

In other words, if NT transmissions of a given word are made, and the probability of detecting each preamble is P_D , then at least M samples will be received on a P_{acq} probability basis. M will be 3, 4, 5, and 5 in the cases of 2-out-of-3, 3-out-of-4, 3-out-of-5, and 4-out-of-5 majority voting, respectively. NT to acquire on a 95 percent probability basis is shown in the last column of Table 6 for different majority voting. It should be noted that in the case of simple majority voting (i.e., 2-out-of-3 and 3-out-of-5) there are no cases where the detector cannot make a decision once the M samples have been collected. However, in the case where using more than simple majority (i.e., 3-out-of-4, 4-out-of-5) there are occasional cases where the required majority states will not be reached even after the required number of samples have been collected. In these cases, the detector must continue to receive samples beyond the M required to reach a decision.

Summary of Results for Basic Data

The following are the main findings for the performance of the BDF based on the results of this section: 1) an MLS receiver which meets only the minimum requirement for decoding the angle function could have (when minimum signal power density exists at the coverage limit) a bit error rate of 2.7×10^{-2} such that the requirement for $P_{UE} \leq 10^{-6}$ will not be met by processing a single data sample without validation; 2) the required bit error rate of 10^{-4} (in order to satisfy the requirement for $P_{UE} \leq 10^{-6}$ at the MLS coverage limit) can be obtained by improving the SNR in the receiver, or by averaging the received data bits of several samples of the same word, or by combination of both techniques; and 3) when the requirement for $P_{UE} \leq 10^{-6}$ at the MLS coverage limit is satisfied, the other requirement for $P_{UE} \leq 10^{-9}$ at critical regions on the approach path should be easily achieved.

INTEGRITY OF THE AUXILIARY DATA FUNCTION

The Auxiliary Data Function is intended to provide airborne systems and flight crew with data necessary to conduct advanced MLS instrument approaches. Though the contents of only four auxiliary data words are currently defined in Reference 1 to provide ground equipment siting information for use in refining airborne position calculations; auxiliary data may also include: a) meteorological information; b) runway status; and c) other supplementary information.

A preliminary organization of these words is defined by ICAO [3] and FAA [4]. Two auxiliary data word formats have been proposed, one for digital* data (Figure 6a) and one for alphanumeric character data (Figure 6b). In either case the word length is 76 bits consisting of 12 bits for a preamble, 8 bits for the address, and 56 bits for information and parity as shown in Figure 6a and b. Since alphanumeric data (i.e., weather information) are not used in critical applications, the performance of validation techniques for these data is not discussed here, but is briefly presented in Reference 6. Unlike the Basic Data Function where a different preamble is used for each word, the Auxiliary Data Function has three different preambles as given by Reference 3, followed by an 8-bit address code to designate different auxiliary data words.

As for the Basic Data Function, validation performance is quantified in terms of the probability of correctly decoding valid data, and the probability of undetected errors remaining in the data. Of the various ways performance could be improved over the baseline, the techniques analyzed here include: 1) using additional airborne processing of multiple samples of a data word; 2) reducing the prevailing bit error probability; and 3) a combination of these two approaches.

Preamble	Address	Data	Parity
'1' '12	'13' '20	'21' '69	'70' '76

(a) Digital Data

Preamble	Address	#1	#2	#3	#4	#5	#6	#7
'1' '12	'13' '20	'21' '28	'29' '36	'37' '44	'45' '52	'53' '60	'61' '68	'69' '76

(b) Alphanumeric Data

FIGURE 6
AUXILIARY DATA WORD ORGANIZATION

Performance of the Address Code of the Auxiliary Data Function

The address code is contained in the 8 bits that follow the preamble of the Auxiliary Data Function and identifies the particular word that is being transmitted [3]. Six of the eight bits encode the actual address and the remaining two bits are parity bits used to check for errors in the six address bits. This enables the encoding of 64 different address numbers (or words). A good parameter for assessing the performance of the address code is the probability of falsely decoding a specific auxiliary data word. Unlike the case for falsely decoded FID's, where the receiver can sometimes reject the following data because it does not match the expected format, even when an address is falsely decoded, the format of the following data is consistent with what the receiver expects.

Figure 7 is used to compute this probability. In this analysis it is assumed that the maximum number of auxiliary data words that would be transmitted at any single site is $L \leq M$ where M is 64. Also, in this figure it is assumed that each word is transmitted at an equal rate, and within a time period T seconds the number of transmissions of each auxiliary data word is N . Figure 7 shows the interactions between the address codes of the Auxiliary Data Function. Within a time period T seconds, each of the L words would be transmitted N times. The auxiliary data address decoder is assumed to correctly recognize which word was transmitted with a probability of P_{D3} . Then within the time period T seconds, $N \cdot P_{D3}$ correct samples can be expected at the output of the decoder for each of the L words. The rest of the samples of each word (i.e., $N(1 - P_{D3})$) will be in error. Part of these incorrect samples is detected and rejected by the address decoder and the other part cannot be detected and will be falsely decoded as one of the other words.

Then the number of falsely decoded samples of word j is given by:

$$N_{FDj} = \frac{P_{UD3}}{M-1} \cdot N \cdot (L-1) \quad (14)$$

where $L \leq M$, $M = 64$, and P_{UD3} is the probability of undetected error within the address code itself. P_{UD3} can be computed using Equation 7 with $n = 8$ and A_j as given in Table 7. This table represents

*The word "digital" data as used by References 3 and 4 is meant to be "numeric" data.

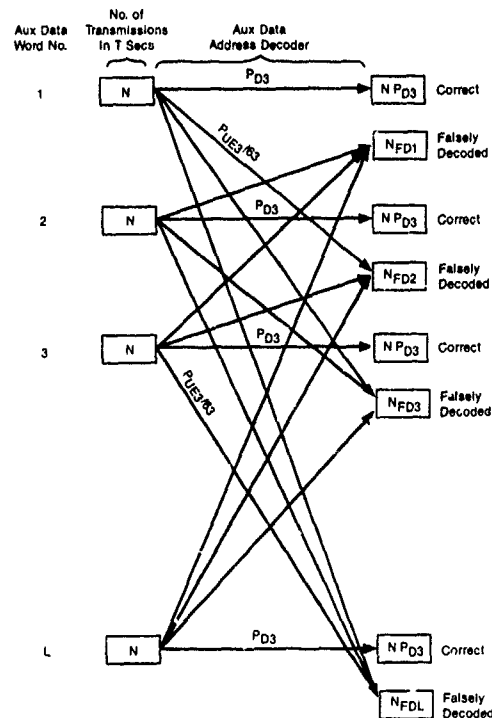


FIGURE 7
SCHEMATIC DIAGRAM SHOWING THE INTERACTIONS
BETWEEN THE ADDRESS CODES OF THE AUXILIARY
DATA FUNCTION

the number of all possible combinations of i errors which are not detected by the parity Equations 15 and 16 of the address code, and is generated by computer using reasoning similar to the one mentioned before. The parity equations of the address code are given as follows [1].

$$C_1 = I_{13} + I_{14} + \dots + I_{19} = 0 \quad (15)$$

$$C_2 = I_{14} + I_{16} + \dots + I_{20} = 0 \quad (16)$$

where "0" in the right hand side of Equations 15 and 16 means "even" and "+" means modulo 2 addition.

Using Equation 14, the probability of falsely decoding the auxiliary data word j can be given by:

$$P_{FDj} = \frac{N_{FDj}}{N} = \frac{L-1}{M-1} \cdot P_{UE3}, \quad j = 1, 2, \dots, L \quad (17)$$

TABLE 7
THE NUMBER OF COMBINATIONS OF i BIT ERRORS NOT DETECTED
BY THE ADDRESS DECODER OF THE AUXILIARY DATA FUNCTION

NUMBER OF BITS HAVING ERRORS i	NUMBER OF COMBINATIONS NOT DETECTED BY PARITY A_i
2	9
3	12
4	19
5	16
6	3
7	4
8	0

It should be noted that the maximum value of L is a function of the repetition rate of each auxiliary data word transmitted through this function and the time slots available for transmission of these words. In Reference 3, the maximum time available for transmission of Auxiliary Data A, B, and C allow a rate of only 21.5 words/sec. If it is assumed that Auxiliary Data A occupy all the available time slots, and the transmission rate for each word within this function is 1 Hz, then $L = 21.5$.

At the MLS coverage limit, $P_e = 2.7 \times 10^{-2}$ and then $P_{UG} = 5.78 \times 10^{-3}$, therefore

$$P_{FDj} = \frac{21.5 - 1}{64 - 1} \times (5.78 \times 10^{-3}) = 1.882 \times 10^{-3}$$

This means that at the MLS coverage limit, the probability of falsely decoding one address code of an auxiliary data function is less than 0.2 percent. This value is very small such that its impact on the auxiliary data is negligible especially if the MLS receiver processor uses an averaging process of multiple samples of the same word (i.e., majority voting).

Performance of the Auxiliary Data Using Hamming Code

The application of the Hamming code to a 64 bit word for single bit error correction and double error detection capability is given in detail in References 8 through 12, and summarized in Reference 6. The parameters of the Hamming code in this application are: 1) the word length $n = 64$; 2) the number of information bits $k = 57$; and 3) the number of parity bits $n - k = 7$. The properties of this code are: 1) it corrects single bit errors; 2) it detects all double bit errors and all odd number of errors; and 3) it does not detect some of the even number of errors, i, starting from four errors ($i = 4, 6, 8, \dots$).

The Hamming code is suitable for this application for a number of reasons, including: (1) it can be readily implemented in hardware for both coding (transmitter) and decoding (receiver) circuits using shift registers and Exclusive OR circuits; (2) it increases the probability of detection; and (3) it decreases the probability of undetected errors.

The probability of detection and the probability of undetected errors for a single error correction and double error detection Hamming code are given in Equations 18 and 19, respectively. It should be noted, however, that P_D is now comprised of two parts: the probability of no errors, plus the probability of a single error which is corrected by the code.

$$P_D = (1 - P_e)^n + nP_e(1 - P_e)^{n-1}, \quad n = 64 \quad (18)$$

$$P_{UE} = \frac{1}{2n} [1 + (1 - 2P_e)^n + 2(n-1)(1 - 2P_e)^{n/2}] - (1 - P_e)^n \quad (19)$$

It should be noted that Equation 19 is the exact formula for P_{UE} as given in Reference 13.

Table 8 shows P_D , P_{err} , P_{UE} , and P_{DE} for four values of the bit error rate (i.e., $P_e = 2.7 \times 10^{-2}$, 10^{-2} , 10^{-3} , and 10^{-4}). From this table it is seen that the requirement for P_{UE} is satisfied when P_e less than 10^{-2} . P_{UE} is also plotted against P_e using Equation 19 and is shown in Figure 8. From this figure it is seen that the bit error rate, P_e , at the MLS coverage limit should be less than 3×10^{-3} to satisfy the requirement for $P_{UE} \leq 10^{-6}$. However, since the prevailing bit error rate for the minimum receiver at maximum range is 2.7×10^{-2} , either additional signal processing would be necessary to meet the requirements or the bit error probability would have to be reduced.

TABLE 8
PERFORMANCE OF (64, 57) HAMMING CODE ON AUXILIARY DATA WORD

P_e	2.7×10^{-2}	10^{-2}	10^{-3}	10^{-4}
P_D	0.481	0.865	0.9981	1.0
P_{err}	0.519	0.135	1.935×10^{-3}	2×10^{-5}
P_{UE}	1.169×10^{-3}	5.765×10^{-5}	9.810×10^{-9}	1.035×10^{-12}
P_{DE}	0.9988	0.9999	1.0	1.0

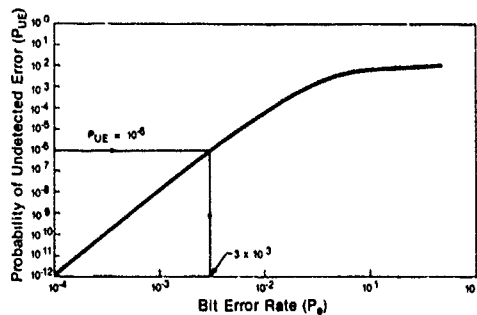


FIGURE 8
PROBABILITY OF UNDETECTED ERROR OF THE (64, 57)
HAMMING CODE VERSUS BIT ERROR RATE

Performance of the Auxiliary Data Using Majority Voting in Addition to Hamming Code

One airborne processing method which could be used with auxiliary data to improve the probability of detection and reduce the probability of an undetected error is the majority voting as explained before and shown in Figure 4. In the same way as it could be used for basic data, the majority voting check N samples of the same bit of a specific word and take the majority state. Once the data has passed through the detector, a Hamming decoder could be used, depending on the established coding convention.

Figure 9 shows the relation between the effective bit error rate, P_{e1} , at the output of the majority voting with the bit error rate, P_e , at its input for 2-out-of-3, 3-out-of-4, 3-out-of-5, and 4-out-of-5 detectors. Using this figure and the requirement for $P_{e1} \leq 3 \times 10^{-3}$ in the case of the Hamming code, the required bit error rate is computed for 2-out-of-3 majority voting, which is adequate in this case, and is shown in Table 9. In this table, column 1 lists the receiver processing, column 2 is the bit error rate to satisfy the requirement of $P_{ue} \leq 10^{-6}$, and column 3 is the improvement in SNR to achieve the required bit error rate given that the bit error rate at the coverage limit is 2.7×10^{-2} . The approximate SNR improvement, needed over the minimum receiver to achieve this bit error rate (if the receiver processes one sample only) is 2.4 dB. This improvement in the SNR is computed using the foot-note equation in the second section for DPSK detector. Column 4 is the number of transmissions required to acquire the data with 95 percent confidence.

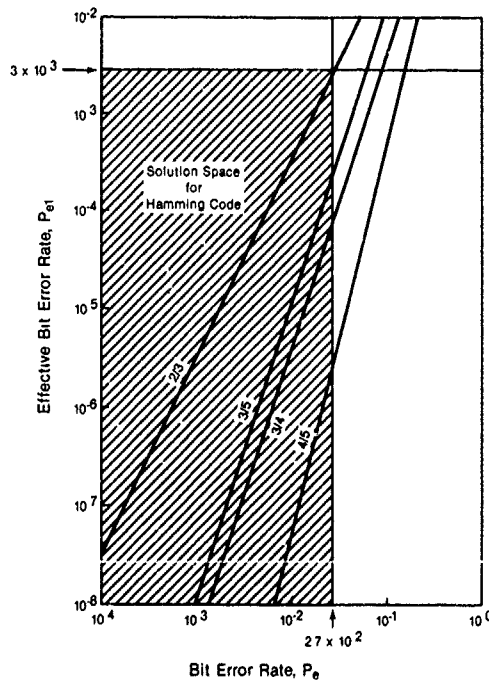


FIGURE 9
THE EFFECTIVE BIT ERROR RATE, P_{e1} , VERSUS THE
BIT ERROR RATE, P_e , FOR AUXILIARY DATA WITH
MAJORITY VOTING

TABLE 9
EXAMPLE TECHNIQUES TO ATTAIN
 $P_{UE} = 10^{-6}$ **FOR MLS AUXILIARY**
DATA AT LIMITS OF COVERAGE⁽¹⁾

RECEIVER PROCESSING		APPROXIMATE BIT ERROR RATE TO ATTAIN $P_{UE} = 1 \times 10^{-6}$	APPROXIMATE SNR IMPROVEMENT NEEDED IN THE MIN. RECEIVER TO ACHIEVE THE BIT ERROR RATES IN COLUMN 2	TRANSMISSIONS REQUIRED TO ACQUIRE AN AUX DATA WORD (0.95 PROBABILITY) [see Reference 6]
Hamming Code (2) (7-Parity Bits)		0.003	2.4 dB	1
Majority Voting (3) Plus Hamming Code	2/3	0.027	0 dB	7

Notes: 1) All values are approximate.
 2) Provided in existing auxiliary data word format (Reference 3).
 3) Majority voting performance out of N voting on each bit prior to the parity check where $m > N/2$.

Summary of Results for Auxiliary Data

Based on the results of this section, the performance analysis of the Auxiliary Data Function indicates the following:

1. An MLS receiver which meets only the minimum requirements could have (when minimum signal power density exists at the coverage limit) a bit error rate high enough that the requirement for P_{UE} will not be met.
2. Improvements in bit error rate can be attained by improving the performance of the demodulator, or by improving the SNR in the receiver, or by averaging the received data bits of several samples of the same word, or any combination of the above methods. Examples of the performance of two decoding/processing combinations and the approximate bit error rates which satisfy the $P_{UE} \leq 10^{-6}$ requirement for the auxiliary data are given in Table 9.
3. Similar to the analysis of Basic Data Function in the third section, when the requirement for $P_{UE} \leq 10^{-6}$ at the minimum SNR is satisfied, the other requirement for $P_{UE} \leq 10^{-9}$ at critical regions (as defined by the area within 10 nmi from the runway) should be easily achieved due mainly to the improvement in SNR exceeding 6 dB.

CONCLUSIONS

Based on the analyses presented in this paper, the following conclusions are given for the integrity of the MLS Basic and Auxiliary Data Functions. It should be noted that some of these conclusions are common for both functions, some apply only to the Basic Data Function, and some apply only to the Auxiliary Data Function.

1. For basic and auxiliary data, an MLS receiver with minimum sensitivity will have a bit error rate that is too large to meet the assumed requirement, $P_{UE} \leq 10^{-6}$.
2. For basic data, the required improvement in the bit error rate can be attained by improving the SNR in the receiver, or by averaging the received data bits of several samples of the same word, or by combination of these techniques. For example, Section three shows that the assumed requirement for $P_{UE} \leq 10^{-6}$ at the MLS coverage limit can be achieved by using one of the following methods:
 - a. Improving the SNR at the receiver by about 4.7 dB. The number of transmissions required in this case is one sample to acquire a basic data word with 95 percent probability.
 - b. Using a 3-out-of-4 majority voting before the decoder without requiring an improvement in the SNR. The number of transmissions required in this case is 8 to acquire a basic data word with 95 percent probability.
 - c. Using a 2-out-of-3 majority voting before the decoder combined with about 1.8 dB improvement in the SNR at the receiver. The number of transmissions required in this case is 4 to acquire a basic data word with 95 percent probability.
3. Similarly, for auxiliary data, Section four shows that the assumed requirement for $P_{UE} \leq 10^{-6}$ at the MLS coverage limit can be achieved by using one of the following methods:
 - a. Improving the SNR at the receiver by about 2.4 dB with (64,57) Hamming code. The number of transmissions in this case is one sample to acquire an auxiliary data word with 95 percent probability.

b. Using a 2-out-of-3 majority voting before the decoder and decoding the (64,57) Hamming code without requiring an improvement in the SNR at the receiver. The number of transmissions in this case is 7 to acquire an auxiliary data word with 95 percent probability.

4. When the requirement for $P_{UE} \leq 10^{-6}$ is satisfied at the MLS coverage limit for basic and auxiliary data, the other requirement for $P_{UE} \leq 10^{-9}$ at critical regions should be easily achieved.

REFERENCES

1. International Civil Aviation Organization, All Weather Operations Panel, Report of the Tenth Meeting and Background Information Paper No. 105, Montreal, Canada, September 1984.
2. International Civil Aviation Organization, All Weather Operations Panel, Report of the Data Function Subgroup, Washington, D.C., June 1984.
3. International Standards and Recommended Practices, Aeronautical Telecommunications, Annex 10 to the Convention on International Civil Aviation, Vol. 1, International Civil Aviation Organization, Montreal, Canada, April 1985.
4. Federal Aviation Administration, "Microwave Landing System (MLS), Interoperability and Performance Requirements," FAA-STD-022c, Washington, D.C., June 1986.
5. P. F. Panter, Modulation, Noise, and Spectral Analysis (Applied to Information Transmission), McGraw-Hill Book Company, New York, NY, 1965.
6. M. B. El-Arini, "Integrity of the MLS Data Functions," Report No. MTR-84W238, The MITRE Corporation, McLean, VA, May 1985.
7. G. W. Flathers, "MLS Basic Data Function Review," Report No. MTR-84W3, The MITRE Corporation, McLean, VA, February 1984.
8. R. W. Hamming, Coding and Information Theory, Prentice-Hall, Inc., Englewood Cliffs, NJ, 1980.
9. R. W. Hamming, "Error Detecting and Error Correcting Codes," Bell Systems Technical Journal, Vol. 29, pp. 147-160, April 1950.
10. S. Lin, An Introduction to Error-Correcting Codes, Prentice-Hall, Inc., Englewood Cliffs, NJ, 1970.
11. W. W. Peterson and E. J. Eldon, Error-Correcting Codes, Second Ed., The MIT Press, Cambridge, MA, 1972.
12. F. J. MacWilliams and N. J. A. Sloane, The Theory of Error-Correcting Codes, North-Holland Publishing Co., Amsterdam, Holland, 1977.
13. S. K. L. Y- Cheong, E. R. Barnes, and D. U. Friedman, "On Some Properties of the Undetected Error Probability of Linear Codes," IEEE Trans. on Information Theory, Vol. IT-25, No. 1, pp. 110-112, January 1979.

PART VII

Special Topics

FAULT DETECTION AND ISOLATION (FDI) TECHNIQUES FOR GUIDANCE AND CONTROL SYSTEMS

by

Mark A. Sturza
Litton Systems, Inc.
Guidance and Control Systems Division
5500 Canoga Avenue
Woodland Hills, California 91367-6698
United States

SUMMARY

Fault Detection and Isolation (FDI) techniques are described with particular emphasis on strapdown inertial system and Global Positioning System (GPS) applications. A generalized measurement model is considered with a single fault, step bias shift fault model. The parity vector is developed and analyzed. Equivalence is established between the parity and innovations approaches to FDI. A fault detection technique based on the parity vector is presented and analyzed.

The Probabilities of False Alarm (PFA) and Missed Detection (PMD) are derived. The Detector Operating Characteristic (DOC) relating these probabilities is constructed. DOCs are presented for several inertial sensor and GPS satellite configurations. A maximum likelihood fault identification technique is described. A nonparity approach to FDI analysis is presented. The final section lists areas for future work.

1. INTRODUCTION

Fault detection and isolation techniques are required in many guidance and control applications. Two applications of topical interest are strapdown inertial systems (INS or AHRSS) with redundant sensors [1] and GPS navigation systems using redundant satellite measurements [2]. Conventional inertial systems use three gyros and three accelerometers mounted on orthogonal axes. Redundant sensor inertial systems typically incorporate four, five or six sets of inertial sensors. These systems are used to provide fail-safe and fault-tolerant operation for flight control and navigation. A fail-safe system detects out of spec sensor performance and takes itself off line. A fault-tolerant system detects and isolates out of spec sensors and continues operating as a fault-tolerant or fail-safe system with the remaining sensors. A minimum of four sensors are required for fail-safe operation and a minimum of five for fault tolerant operation.

Conventional GPS navigation requires measurement from four satellites to resolve three spatial dimensions and time. Typically five or six satellites are available. By using the redundant measurements it is possible to provide integrity monitoring. This is a requirement if GPS is to be certified on a sole means navigation system [3]. A minimum of five satellite measurements are required to detect an out of spec satellite. If measurements from six or more satellites are available it is possible to isolate the failed satellites and continue operating with the remaining ones.

In both the inertial system and the GPS application measurement FDI techniques are required. Section 2 describes the measurement model used in subsequent analysis. The parity vector is developed in Section 3. Sections 4 and 6 describe parity based fault detection and isolation techniques. Examples of fault detection performance are presented in Section 5. Section 7 describes an intuitive, nonparity approach to FDI. Topics for future work are discussed in Section 8.

2. MEASUREMENT MODEL

The general case of M measurements in an N dimensional state space, $M \geq N + 1$, is addressed. The measurement equation is:

$$\underline{z} = H \underline{x} + \underline{n} + \underline{b} \quad (2-1)$$

where

\underline{z} is the $M \times 1$ vector of measurements compensated with all a priori information.

H is the $M \times N$ measurement matrix which transforms from the state space to the measurement space, $\text{rank } [H] = N$;

\underline{x} is the $N \times 1$ state vector;

\underline{n} is the $M \times 1$ vector of Gaussian measurement noise, $E[\underline{n}] = \underline{0}$ and $\text{COV}[\underline{n}] = \sigma_n^2 I_M$;

\underline{b} is the $M \times 1$ vector of uncompensated measurement biases (faults).

The fault model is a single measurement source failure resulting in a step bias shift. A failure of measurement source i is modeled by $\underline{b} = \underline{b}_i$ where \underline{b}_i is a $M \times 1$ vector with i th element B and zeros elsewhere. If there are no faults then $\underline{b} = \underline{0}$.

In the inertial system application the gyro and accelerometer measurements are considered separately. In both cases $N = 3$ and $M = 4, 5$, or 6 . The measurement vector, \underline{z} , contains the $\Delta\theta$ s (gyro measurements) or ΔV s (accelerometer measurements) compensated for scale factor and biases. The measurement matrix, H , consists of the unit projection vectors from the sensor axis to the orthogonal body axis. It is assumed that H is known perfectly (there are no misalignment errors).

In the GPS application the pseudorange residual ($\tilde{P}\tilde{R}$) and delta-range residual ($\tilde{\Delta}\tilde{R}$) measurements are considered separately. For both types of measurements $N = 4$ and $M = 5$ or 6 . The measurement vector, \underline{z} , contains the $\tilde{P}\tilde{R}$ measurements or the $\tilde{\Delta}\tilde{R}$ measurements compensated for atmospheric effects and clock biases. The $\tilde{P}\tilde{R}$ measurement matrix consists of the line-of-sight (LOS) vectors to the satellites with 1s in the fourth column corresponding to the clock bias state. The $\tilde{\Delta}\tilde{R}$ measurement matrix consists of the difference of the LOS vectors over the delta range interval with 1s in the fourth column corresponding to the clock bias rate state. It is again assumed that H is known perfectly.

3. PARITY VECTOR

In this section a $M - N \times 1$ parity vector, \underline{p} , is constructed. It is shown that:

1. \underline{p} is independent of the state vector \underline{x}
2. if there is no fault, $\underline{b} = \underline{0}$, then $E[\underline{p}] = \underline{0}$
3. if there is a fault then $E[\underline{p}]$ is a function of B .

Thus \underline{p} can be used for fault detection. An $M \times 1$ fault vector, \underline{f} , is constructed by transforming \underline{p} to the measurement space. It is shown that \underline{f} is identical to the measurement innovations vector, $\tilde{\underline{z}}$. Thus the parity and innovations approaches to FDI are identical.

The least squares estimate of the state vector in (2-1) is

$$\begin{aligned}\hat{\underline{x}} &= H^* \underline{z} \\ &= \underline{x} + H^* (\underline{n} + \underline{b})\end{aligned}\quad (3-1)$$

where

$H^* = (H^T H)^{-1} H^T$ in the $N \times M$ generalized inverse of H . The matrix H^* represents a transformation from the measurement space to the state space.

For a given $M \times N$ measurement matrix H with rank N it is possible to find a $(M - N) \times M$ matrix P such that rank $[P] = M - N$, $P P^T = I_{M-N}$, and $P H = 0$. The matrix P spans the null space of H (the parity space) and $P H^* T = 0$. The rows of P are orthogonal unit basis vectors for the parity space.

The $M \times M$ matrix

$$A = \begin{bmatrix} H^* \\ P \end{bmatrix}\quad (3-2)$$

has rank M . The linear transformation represented by A separates the M dimensional measurement space into two sub-spaces - an N dimensional state space and an $M - N$ dimensional parity space:

$$\begin{array}{c} \uparrow N \\ \downarrow M-N \\ \hline \begin{bmatrix} \text{STATE SPACE} \\ \text{PARITY SPACE} \end{bmatrix} \end{array} = A \begin{bmatrix} \text{MEASUREMENT SPACE} \end{bmatrix} \begin{array}{c} \uparrow M \\ \downarrow \end{array}\quad (3-3)$$

The $(M - N) \times 1$ parity vector given by

$$\begin{aligned}\underline{p} &= \underline{Pz} \\ &= \underline{P}(\underline{n} + \underline{b})\end{aligned}\quad (3-4)$$

is independent of \underline{x} since $\underline{PH} = 0$ (if H is not exact then the parity vector is no longer independent of the state). The elements of \underline{p} are jointly normally, characterized by their expected value and covariance:

$$E[\underline{p}] = \underline{Pb} \quad (3-5)$$

$$\begin{aligned}\text{COV}[\underline{p}] &= E[(\underline{p} - E[\underline{p}])(\underline{p} - E[\underline{p}])^T] \\ &= E[\underline{Pnn}^T \underline{P}^T] \\ &= \sigma_n^2 \underline{I}_{M-N}.\end{aligned}\quad (3-6)$$

Thus the elements of \underline{p} are uncorrelated with equal variance, the same variance as the measurement noise. If there is no fault, $\underline{b} = \underline{0}$, then $E[\underline{p}] = 0$.

The $M \times M$ matrix

$$\underline{A}^{-1} = [\underline{H} \mid \underline{P}^T] \quad (3-7)$$

has rank M and represents the inverse transformation represented by \underline{A} . It transforms from the state space and the parity space to the measurement space.

$$\begin{array}{c} \updownarrow \\ M \\ \downarrow \end{array} \left[\begin{array}{c} \text{MEASUREMENT SPACE} \end{array} \right] = \underline{A}^{-1} \left[\begin{array}{c} \text{STATE SPACE} \\ \hline \text{PARITY SPACE} \end{array} \right] \begin{array}{c} \updownarrow \\ N \\ \downarrow \\ \updownarrow \\ M-N \\ \downarrow \end{array} \quad (3-8)$$

Thus the transformation of the parity vector to the measurement space is given by

$$\begin{aligned}\underline{f} &= [\underline{H} \mid \underline{P}^T] \begin{bmatrix} 0 \\ \underline{p} \end{bmatrix} \\ &= \underline{P}^T \underline{p} \\ &= \underline{P}^T \underline{P} \underline{z} \\ &= \underline{S} \underline{z}\end{aligned}\quad (3-9)$$

where $\underline{S} = \underline{P}^T \underline{P}$ is an $M \times M$ matrix. The matrix \underline{S} has rank $M - N$ and is idempotent ($\underline{S}^2 = \underline{S}$). The $M \times 1$ fault vector, \underline{f} , is characterized by

$$E[\underline{f}] = \underline{Sb} \quad (3-10)$$

and

$$\text{COV}[\underline{f}] = \sigma_n^2 \underline{S}. \quad (3-11)$$

The measurement innovation is the difference between the actual measurement and the best estimate of the measurement based on the estimated state:

$$\tilde{z} = z - \hat{z} \quad (3-12)$$

where

\tilde{z} is the $M \times 1$ vector of measurement innovations

\hat{z} is the $M \times 1$ vector of estimated measurements

The best estimate of the z based on the estimated state is:

$$\begin{aligned} \hat{z} &= E[z | \hat{x}] \\ &= H\hat{x} \\ &= HH^*z \end{aligned} \quad (3-13)$$

Thus

$$\begin{aligned} \tilde{z} &= z - \hat{z} \\ &= (I_M - HH^*)z. \end{aligned} \quad (3-14)$$

It is shown in the appendix that $S = I_M - HH^*$, so the measurement innovations vector, \tilde{z} , and the fault vector, f , are identical.

4. FAULT DETECTION

Fault detection techniques are based on hypothesis testing. A decision variable, D , is constructed and tested against a threshold, T . The hypothesis test is

$$D \underset{H_0}{\overset{H_1}{>}} T \quad (4-1)$$

where H_0 is the null hypothesis (no fault, $\underline{b} = \underline{0}$) and H_1 is the fault hypothesis ($\underline{b} = \underline{b}_i$ for some i). The performance of the test is characterized by the probability of false alarm (P_{FA}) and the probability of missed detection (P_{MD}):

$$P_{FA} = P[D > T | H_0] \quad (4-2)$$

$$P_{MD} = P[D < T | H_1]. \quad (4-3)$$

The detector operating characteristics, (DOCs), are obtained by plotting P_{MD} vs P_{FA} for various combinations of parameters.

The decision variable considered is the square of the magnitude of the parity vector which is equivalent to the quadratic form

$$\begin{aligned} D &= \underline{p}^T \underline{p} \\ &= (\underline{n} + \underline{b})^T S (\underline{n} + \underline{b}). \end{aligned} \quad (4-4)$$

It is equal to the square of the magnitude of the fault vector (and hence the measurement innovations vector):

$$\begin{aligned} \underline{f}^T \underline{f} &= (\underline{n} + \underline{b})^T S^T S (\underline{n} + \underline{b}) \\ &= (\underline{n} + \underline{b})^T S (\underline{n} + \underline{b}) \\ &= D. \end{aligned} \quad (4-5)$$

The hypothesis test is characterized by

$$P_{FA} = P[D > T \mid \underline{b} = \underline{0}] \quad (4-6)$$

and

$$P_{MD} = P[D < T \mid \underline{b} = \underline{b}_i \text{ for some } i]. \quad (4-7)$$

Assuming that measurement faults are equally likely

$$P_{MD} = \frac{1}{M} \sum_{i=1}^M P[D < T \mid \underline{b} = \underline{b}_i]. \quad (4-8)$$

If $\underline{b} = \underline{0}$ then $E[\underline{f}] = \underline{0}$ and D/σ_n^2 has chi-square distribution which $M - N$ degrees of freedom. Thus

$$P_{FA} = Q(T/\sigma_n^2 \mid M - N) \quad (4-9)$$

where $Q(\chi^2 \mid r) = 1 - P(\chi^2 \mid r)$ and $P(\chi^2 \mid r)$ is the chi-square probability function

$$P(\chi^2 \mid r) = \left[2^{r/2} \Gamma\left(\frac{r}{2}\right) \right]^{-1} \int_0^{\chi^2} t^{\frac{r}{2}-1} e^{-\frac{t}{2}} dt. \quad (4-10)$$

P_{FA} depends on $M - N$, the number of redundant measurements, and is otherwise independent of H . For given P_{FA} , $M - N$, and σ_n^2 the required threshold is given by

$$T(P_{FA}, M - N, \sigma_n^2) = \sigma_n^2 Q^{-1}(P_{FA} \mid M - N) \quad (4-11)$$

where $Q^{-1}(P \mid r)$ is the inverse function of $Q(\chi^2 \mid r)$. If $\underline{b} = \underline{b}_i$ then $E[\underline{p}] = P\underline{b}_i$ and D/σ_n^2 has noncentral chi-square distribution with $M - N$ degrees of freedom and noncentrality parameter

$$\begin{aligned} \theta_i &= \left(B^2 / \sigma_n^2 \right) \sum_{j=1}^{M-N} P_{ji}^2 \\ &= \left(B^2 / \sigma_n^2 \right) S_{ii} \end{aligned} \quad (4-12)$$

VIII-6

Thus

$$P_{MD} | b_i = P \left(T/\sigma_n^2 \mid M-N, \theta_i \right) \quad (4-13)$$

and

$$P_{MD} = \frac{1}{M} \sum_{i=1}^M P \left(T/\sigma_n^2 \mid M-N, \frac{B^2}{\sigma_n^2} S_{ii} \right) \quad (4-14)$$

where $P(\chi^2 | r, \theta)$ is the noncentral chi-square probability function

$$P(\chi^2 | r, \theta) = \sum_{j=0}^{\infty} \frac{\theta^j}{j!} e^{-\theta/2} P(\chi^2 | r+2j). \quad (4-15)$$

The detector operating characteristic is given by

$$P_{MD} \left(P_{FA}, H, \frac{B^2}{\sigma_n^2} \right) = \frac{1}{M} \sum_{i=1}^M P \left(Q^{-1}(P_{FA}) \mid M-N \mid M-N, \frac{B^2}{\sigma_n^2} S_{ii} \right) \quad (4-16)$$

P_{MD} depends only on the ratio B^2/σ_n^2 and not on the individual values.

5. DETECTOR OPERATING CHARACTERISTICS (DOCs)

DOCs are presented for two 4 and two 6 sensor inertial system configurations, and for 6 GPS satellites at two instants of time. The first inertial system configuration is four sensors arranged three orthogonal and one on the diagonal. Thus

$$M = 4 \quad N = 3$$

$$H = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \\ \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} \end{bmatrix} \quad P = \begin{bmatrix} \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{2}} \end{bmatrix}$$

$$S = \begin{bmatrix} \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{-1}{\sqrt{12}} \\ \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{-1}{\sqrt{12}} \\ \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{-1}{\sqrt{12}} \\ \frac{-1}{\sqrt{12}} & \frac{-1}{\sqrt{12}} & \frac{-1}{\sqrt{12}} & \frac{1}{2} \end{bmatrix}$$

DOCs for this configuration are shown in Figure 1 for $B/\sigma_n = 10, 15, 20$, and 25 . The probability of detecting a $250 \mu\text{g}$ bias shift of an accelerometer with $10 \mu\text{g}$ of noise ($B/\sigma_n = 25$) at a P_{FA} of 10^{-6} is 0.99999999 ($1 - P_{MD}$). The probability of detecting a $150 \mu\text{g}$ bias shift ($B/\sigma_n = 15$) of the same accelerometer with the same P_{FA} is 0.9 . The probability of detecting an $0.03^\circ/\text{hr}$ bias shift of a gyro with $0.0015^\circ/\text{hr}$ noise ($B/\sigma_n = 20$) at a P_{FA} of 10^{-6} is 0.9997 . The second 4 sensor inertial configuration is 4 sensors equally spaced on a cone with half angle of 54.736 degrees. The DOCs for this configuration are shown in Figure 2 for $B/\sigma_n = 7.5, 10, 15$, and 20 . Figure 3 is a comparison of DOCs from both 4 sensor configurations. It shows that for a given B/σ_n the 4 sensors on a cone ($\alpha = 54.736^\circ$) provides significantly superior fault detection capability.

The two 6 sensor inertial system configurations considered are: skewed triads (6 sensors equally spaced on a cone with half angle of 54.736 degrees) and dodecahedral (6 sensors equally spaced on a cone with a half angle of 63.435 degrees). The DOCs for these two configurations are identical (the diagonals of the S matrix for both configurations are identical). The DOCs are shown in Figure 4 for $B/\sigma_n = 7.5, 10, 12.5$, and 15 . It is apparent that the 6 sensor configurations provide better fault detection capability than the 4 sensor configurations. For a $P_{MD} = 10^{-7}$ and a $P_{FA} = 10^{-6}$ a B/σ_n ratio of 20 is required for the 4 sensors on a cone ($\alpha = 54.736^\circ$) configuration while only 15 is required for the 6 sensor configurations.

The GPS examples are based on the current test constellation. The measurement geometry resulting from satellites 3, 6, 9, 11, 12, and 13 in the Los Angeles area ($N34^\circ 17'$, $W118^\circ 51'$) on 20 January 1987 at 1100 and 1130 GMT is considered. In the GPS application the measurement geometry changes as a function of time (as the satellites and earth move). The DOCs are shown in Figures 5 and 6 for 1100 and 1130 GMT respectively. A drastic change in detection performance occurs between the two times. At 1100 a B/σ_n ratio of 20 and a $P_{FA} = 10^{-5}$ results in a P_{MD} of 10^{-6} while at 1130 the resulting P_{MD} is 10^{-1} , a change of five orders of magnitude in 30 minutes!

The GPS $\ddot{P}\ddot{R}$ measurement noise depends on the class of service used. The values are 1.5 meters for the PPS (military service) and 15 meters for the SPS (civil service). The $\ddot{D}\ddot{R}$ measurement noise is 0.02 meters/second. The following analysis is based on the DOCs presented in Figure 5. The probability of detecting a $\ddot{P}\ddot{R}$ bias shift such that $B/\sigma_n = 20$ (30 meters PPS, 300 meters SPS) with a $P_{FA} = 10^{-6}$ is 0.999999 . The same probabilities would hold for a $\ddot{D}\ddot{R}$ bias shift of 0.4 meters/second.

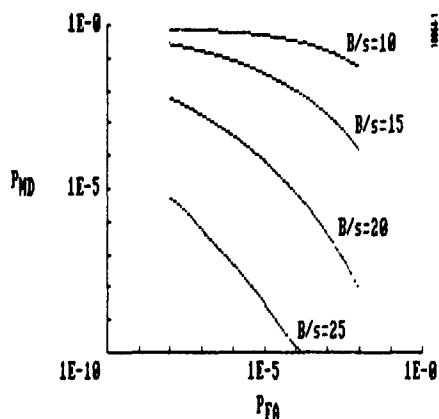


Figure 1. Four Sensors: Three Orthogonal and One on the Diagonal

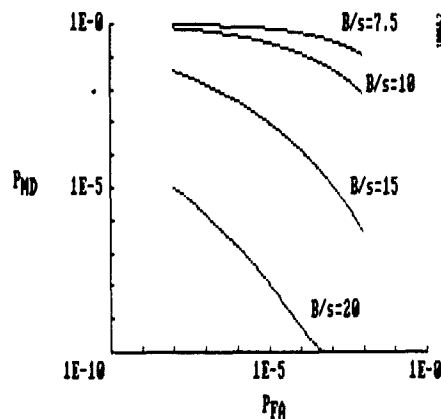


Figure 2. Four Sensors on Cone ($\alpha = 54.736$ Degrees)

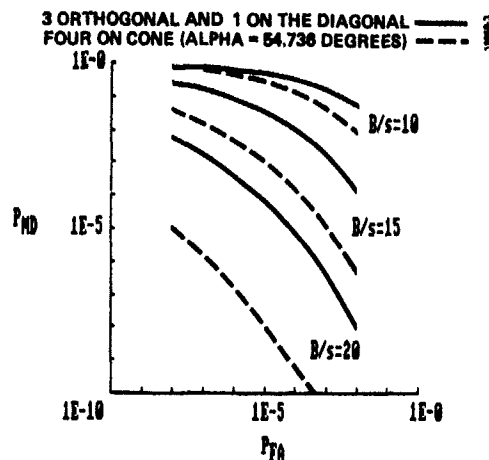


Figure 3. Comparison of Four Sensor Configurations

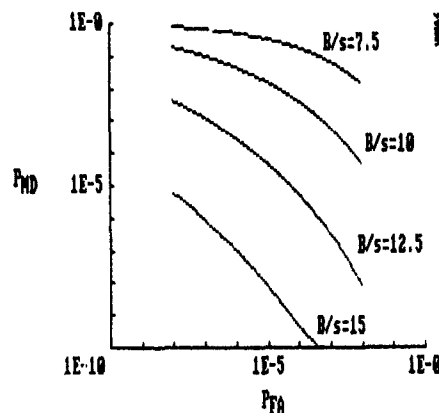


Figure 4. Two Skewed Triads
($\alpha = 54.736$ Degrees)
Dodecahedral
($\alpha = 63.435$ Degrees)

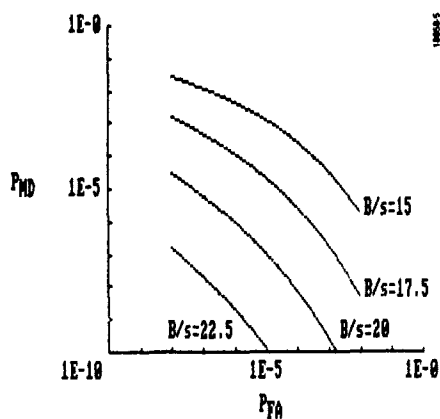


Figure 5. GPS Satellites 3, 6, 9, 11, 12, 13
N34 17, W118 51 1100 GMT
20 January 1987

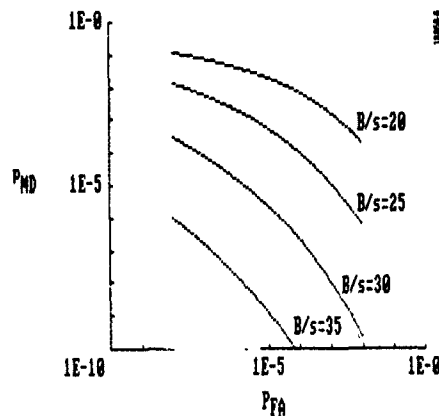


Figure 6. GPS Satellites 3, 6, 9, 11, 12, 13
N34 17, W118 51 1130 GMT
20 January 1987

6. FAULT ISOLATION

A Maximum Likelihood Estimation (MLE) approach to fault isolation is presented [4]. If a fault has occurred then $\underline{b} = \underline{b}_i$ for some i . The goal of fault isolation is to identify which i , which measurement source is at fault. The MLE approach is to determine the i that maximizes.

$$P(\underline{p} | \underline{b}_i). \quad (6-1)$$

This is equivalent to maximum a posterior (MAP) estimation (where i is chosen to maximize $P(\underline{b}_i | \underline{p})$) provided that $P(\underline{b}_i)$ is independent of i , i.e., all measurement source failures are equally likely.

The parity vector has multivariate normal density with mean \underline{Pb} and covariance $\sigma_n^2 \underline{I}_{M-N}$. Thus

$$P(\underline{p} | \underline{b}_i) = \frac{1}{(2\pi\sigma_n^2)^{\frac{M-N}{2}}} \exp \left[-\frac{1}{2\sigma_n^2} (\underline{p} - \underline{Pb}_i)^T (\underline{p} - \underline{Pb}_i) \right]. \quad (6-2)$$

The i that maximizes $P(\underline{p} | \underline{b}_i)$ will maximize any monotonically increasing function of $P(\underline{p} | \underline{b}_i)$. Thus the i that maximizes $P(\underline{p} | \underline{b}_i)$ will maximize

$$\begin{aligned} & -(\underline{p} - \underline{Pb}_i)^T (\underline{p} - \underline{Pb}_i) \\ &= -\underline{p}^T \underline{p} + 2 \underline{b}_i^T \underline{P}^T \underline{p} - \underline{b}_i^T \underline{P}^T \underline{P} \underline{Pb}_i. \end{aligned} \quad (6-3)$$

Since $-\underline{p}^T \underline{p}$ is independent of i it can be dropped. Taking advantage of the structure of \underline{b}_i , and that $\underline{p} = \underline{Pz}$ gives

$$\text{MAX}_i \left\{ 2 B S_i z - B^2 S_{ii} \right\} \quad (6-4)$$

where S_i is the i th row of S . The value of B that maximizes (6-4) is obtained by setting the derivative of (6-4) with respect to B to zero. This gives

$$\hat{B} = \frac{S_i z}{S_{ii}}. \quad (6-5)$$

Substituting back into (6-4) and noting that $f_i = S_i z$ gives

$$\text{MAX}_i \left\{ \frac{f_i^2}{S_{ii}} \right\}. \quad (6-6)$$

Thus the MLE fault identification technique is to select the i that maximizes $\frac{f_i^2}{S_{ii}}$ as the faulty measurement source.

The probability of misidentification, assuming measurement source i is faulty, is

$$P_{MI} | b_i = P \left[\frac{f_i^2}{S_{ii}} < \text{MAX}_j \left\{ \frac{f_j^2}{S_{jj}} \right\} | b_i \right]. \quad (6-7)$$

Assuming that measurement faults are equally likely

$$P_{MI} = \frac{1}{M} \sum_{i=1}^M P \left[\frac{f_i^2}{S_{ii}} < \text{MAX}_j \left\{ \frac{f_j^2}{S_{jj}} \right\} | b_i \right] \quad (6-8)$$

7. NONPARITY MEASUREMENT FDI

This section describes an intuitive approach to FDI analysis. Individual measurements are deleted from the measurement equation (2-1), to form M reduced order measurement equations. The reduced order equation deleting measurement j is:

$$R_j \underline{z} = R_j H \underline{x} + R_j \underline{n} + R_j \underline{b} \quad (7-1)$$

where R_j is the identity matrix I_M with the j th diagonal element zeroed. The reduced order state estimates are:

$$\begin{aligned} \hat{\underline{x}}_j &= H_j^* \underline{z}; & j &= 1, \dots, M \\ &= \underline{x} + H_j^* (\underline{n} + \underline{b}) \end{aligned} \quad (7-2)$$

where

$$H_j^* = (H^T R_j H)^{-1} H^T R_j. \quad (7-3)$$

It is assumed that H_j^* exists for every j , i.e., that $\text{rank}(H^T R_j H) = N$ for all j . If measurement source i is faulty, $\underline{b} = \underline{b}_i$, then $\hat{\underline{x}}_i$ is independent of \underline{B} and the other $M-1$ estimates are functions of \underline{B} .

Let

$$\begin{aligned} \underline{D}_{ij} &= \hat{\underline{x}}_i - \hat{\underline{x}}_j; & i &\neq j \\ &= (H_i^* - H_j^*) \underline{z} \\ &= (H_i^* - H_j^*) (\underline{n} + \underline{b}). \end{aligned} \quad (7-4)$$

There are $M(M-1)$ \underline{D}_{ij} s each independent of the state vector, \underline{x} . The \underline{D}_{ij} s can be computed by first computing the $\hat{\underline{x}}_k$ s and then differencing, or directly from \underline{z} by multiplying by $(H_i^* - H_j^*)$.

Let

$$d_{ij} = |\underline{D}_{ij}|, \text{ then } d_{ij} = d_{ji} \text{ and there are } \frac{M(M-1)}{2} \text{ unique } d_{ij}\text{s. Assume}$$

1. Measurement source i is faulty, $\underline{b} = \underline{b}_i$
2. The square of the bias shift is much greater than the measurement variance, $B^2 \gg \sigma_n^2$
3. The reduced order measurement geometry provides reasonable observability, $\text{TRACE}[(H^T R_j H)^{-1}]$ is on the same order as $\text{TRACE}[(H^T H)^{-1}]$.

Then intuitively it is expected that:

- $d_{jk} \ll B$ if $j \neq i$ and $k \neq i$ since both $\hat{\underline{x}}_j$ and $\hat{\underline{x}}_k$ are functions of \underline{B} .
- $d_{jk} = \frac{N}{M} B$ if $j = i$ or $k = i$ since only one of $\hat{\underline{x}}_j, \hat{\underline{x}}_k$ is a function of \underline{B}

FDI techniques can be developed based on the d_{ij} . A fault detection technique is to use hypothesis testing with a decision variable:

$$D = \sum_{i,j} d_{ij}. \quad (7-5)$$

A combined fault detection and identification technique is a set of M hypothesis tests, one for each measurement source. The test for source i is:

$$D_i \underset{H_0}{\overset{H_i}{\gtrless}} T \quad (7-6)$$

where

$$D_i = \sum_{j=1 \text{ or } k=i} D_{ij}$$

$$H_0: \underline{b} \neq \underline{b}_i \quad (7-7)$$

$$H_i: \underline{b} = \underline{b}_i.$$

8. FUTURE WORK

Several areas for future work are apparent:

1. The measurement model in Section 2 does not allow for uncompensated non-fault measurement biases. These will be present in all real world applications.
2. The fault model in Section 2 is very limited. More realistic models might include multiple faults and measurement source failures resulting in ramp and/or oscillating bias shifts.
3. In the GPS application the $\tilde{P}\tilde{R}$ and $\tilde{D}\tilde{R}$ measurements are related. It may be possible to take advantage of this relationship to improve FDI performance.
4. The measurement matrix, H , may not be known perfectly. This would result from uncompensated misalignments in the inertial system application and state vector errors in the GPS application. Errors in H result in the parity vector being a function of the state vector. What is the effect on FDI performance? Reference [5] suggests an approach to compensate for an imperfectly known measurement matrix.
5. The DOCs presented in Section 5 show that the four sensors on a cone ($\alpha = 54.736^\circ$) provides significantly better fault detection capability than the three orthogonal and one on the diagonal. What is the best four sensor configurations? Five sensors? Six sensors?
6. The GPS DOCs presented in Section 5 are based on the limited test constellation at one location. What are the average DOCs for the operational constellation over the earth's surface and time?
7. How can P_{MI} defined in Section 6 be calculated?
8. What are the P_{FA} , P_{MD} and P_{MI} for the non-parity measurement FDI technique presented in Section 7?
9. The FDI techniques described in Sections 4, 6, and 7 are based on a single sample. What are the appropriate techniques and performances if multiple samples are used?

These are a few areas that may prove of interest for future work.

APPENDIX

Theorem: $S = I_M - HH^*$

Proof: That $A^{-1} = [H|P^T]$ is the right inverse of

$$A = \begin{bmatrix} H^* \\ P \end{bmatrix}$$

is verified by direct multiplication:

$$AA^{-1} = \begin{bmatrix} H^* \\ P \end{bmatrix} [H|P^T]$$

$$= \begin{bmatrix} H^*H & H^*P^T \\ PH & PP^T \end{bmatrix}$$

$$= \begin{bmatrix} I_N & O \\ O & I_{M-N} \end{bmatrix}$$

$$= I_M.$$

Since A is square and of full rank its left inverse must equal its right inverse. Thus

$$I_M = A^{-1}A$$

$$= [H|P^T] \begin{bmatrix} H^* \\ P \end{bmatrix}$$

$$= HH^* + P^T P$$

$$= HH^* + S$$

Hence

$$S = I_M - HH^*$$

Q.E.D.

REFERENCES

- [1] Schley, W.R.; "The Use of Skewed Inertial Sensors in Flight Control Systems;" SAE Aerospace Technology Conference and Exposition; Long Beach, CA; October, 1986.
- [2] Parkinson, B.W. and Axelrod, P; "A Basis for the Development of Operational Algorithms for Simplified GPS Integrity Checking;" ION Satellite Division Technical Meeting; September 1987.
- [3] 1986 Federal Radio Navigation Plan; DOD-4650.4; DOT-TSC-RSPA-87-3.
- [4] Duda, R.O. and Hart, E.H.; Pattern Clarification and Scene Analysis; New York: John Wiley & Sons, Inc.; 1973; Chapters 2 and 3.
- [5] Hall, S.R. et. al.; "Inflight Parity Vector Compensation for FDI;" IEEE Paper 0547-3578/82/0000-0380; 1982.

ACKNOWLEDGEMENTS

The author wishes to thank Ms. Janet L. King, a Caltech student, for generating the DOCs presented in Section 5.

CONTROL AND ESTIMATION FOR AEROSPACE APPLICATIONS WITH SYSTEM TIME DELAYS

by

Edward J. Knobbe, Ph.D.
OFFICE OF THE CHIEF SCIENTIST
NORTHROP ELECTRONICS DIVISION
2301 W. 120th Street
Hawthorne, CA 90250
USA

1. INTRODUCTION

In many practical aerospace applications, guidance and control accuracy is significantly degraded as a result of system transport lags (or time delays). Historically, solutions to problems of this type were required when significant digital computer data delays were encountered or when human operators (e.g., aircraft pilots) became an integral part of the closed-loop operation. More recently, this type of problem is encountered in current SDI (Strategic Defense Initiative) related problems such as Laser Beam Pointing & Tracking and, Atmospheric Laser Beam Wavefront Correction.

In this chapter, we address the problem of controlling a linear discrete-time system where both the measured output and the process evolution are functions of time-delayed states. The optimal control solution is developed by re-casting the original system representation, with explicit time-delayed states, into a standard regulator form using state vector augmentation. Practical considerations are discussed regarding the implementation of this control law.

This solution was developed by Northrop in the study of Precision Pointing & Tracking of laser beams; it should prove useful in this and other advanced aerospace applications where system time delays are present and where precision guidance and control is required.

2. PROBLEM STATEMENT

The general system representation, with explicit process and measurement (observation) time delays, is given by:

$$X(k+1) = \sum_{i=1}^q A_i(k)X(k+1-i) + B(k)U(k) + W_1(k) \quad (1)$$

$$Y(k) = \sum_{j=1}^p C_j(k)X(k+1-j) + W_2(k) \quad (2)$$

where p, q are integers ≥ 1 ; W_1 and W_2 are zero mean white noise sequences such that

$$\left. \begin{aligned} E \{ W_1(k) W_1^T(k) \} &= V_1(k) \\ E \{ W_2(k) W_2^T(k) \} &= V_2(k) \\ E \{ W_1(i) W_2^T(j) \} &= 0 \end{aligned} \right\} \quad \forall i, j, k$$

Y is an $m \times 1$ observation vector; X is an $n \times 1$ random state vector whose initial uncertainty is uncorrelated with W_1 and W_2 and with initial covariance, Q_0 ; and, U is the control input vector. The objective is to find the control function (functional), $U(k)$ for $k = 1, 2, \dots$, which minimizes an expected quadratic cost function for the linear stochastic regulator defined by Equations (1) and (2). Since linear stochastic tracking problems can be formulated as linear stochastic regulator problems by combining the reference and plant models in an augmented system [1], the system representation defined by Equations (1) and (2) applies equally to the tracking problem.

3. CONTROL DEVELOPMENT

The system representation defined by Equations (1) and (2) can be cast in standard stochastic regulator form by augmenting the state vector with the time-delayed states, i.e.,

$$\bar{X}(k+1) = \bar{A}(k) \bar{X}(k) + \bar{B}(k)U(k) + \bar{W}_1(k) \quad (3)$$

AND

$$Y(k) = \bar{C}(k)\bar{X}(k) + W_2(k) \quad (4)$$

WHERE

$$\bar{A}(k) \triangleq \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & \cdots & 0 \\ 0 & \cdots & 0 & \cdots & 0 \\ A_q(k) & \cdots & \cdots & \cdots & A_1(k) \end{bmatrix}; \quad \bar{X}(k) \triangleq \begin{bmatrix} X(k-q-1) \\ X(k-q-2) \\ \vdots \\ X(k-1) \\ X(k) \end{bmatrix};$$

$$\bar{B}(k) \triangleq \begin{bmatrix} 0 \\ \vdots \\ 0 \\ B(k) \end{bmatrix}; \quad \bar{W}_1(k) \triangleq \begin{bmatrix} 0 \\ \vdots \\ 0 \\ W_1(k) \end{bmatrix}; \quad \bar{C}(k) \triangleq \begin{bmatrix} C_{q-1}(k) \\ C_{q-2}(k) \\ \vdots \\ C_2(k) \\ C_1(k) \end{bmatrix}^T.$$

In equation (3) the control remains the same as in Equation (1) since, physically, past states cannot be changed or controlled. In Equation (4), the observation, Y , and the observation noise, W_2 , also remain unchanged.

For simplicity of illustration, we let $p = q$ in Equations (3) and (4). However, if $q > p$, then Equation (3) remains the same and the appropriate submatrices in \bar{C} are set to zero; i.e., the measurement is not a linear combination of all the delayed states contained in \bar{X} . If $q < p$, then Equation (4) remains the same but the augmented state vector, \bar{X} , must now contain all p delayed states in the measurement. The augmented A matrix will contain zeros for the submatrices, A_1 , which correspond to those integers $> q-1$ but $\leq p-1$. The dimensions of \bar{B} and \bar{W} must also be consistent with the integer p , rather than q . Therefore, without loss of generality, the system defined by Equations (1) and (2) can be represented by the augmented system model described by Equations (3) and (4).

Given that the control is to minimize the expected value of a quadratic cost function of the form,

$$E \left\{ \sum_{k=k_0}^{k_1-1} \left[\bar{X}(k)^T \bar{R}_1(k) \bar{X}(k) + U(k)^T R_2(k) U(k) \right] + \bar{X}(k_1)^T \bar{P}(k_1) \bar{X}(k_1) \right\}, \quad (5)$$

then, the Separation Principle applies [1]. And, the optimal linear stochastic control of the augmented system is given by a deterministic optimal linear controller with state input $\hat{\bar{X}}$, (or estimated state feedback) which is provided by an optimal one-step predictor using the augmented model. That is:

$$U(k) = -\bar{F}(k) \hat{\bar{X}}(k); \quad k = k_0, k_0+1, \dots, k_1 \quad (6)$$

where the control gain, \bar{F} , satisfies the following equation,

$$\bar{F}(k) = \left\{ R_2(k) + \bar{B}(k) \left[\bar{R}_1(k+1) + \bar{P}(k+1) \bar{B}(k) \right] \bar{B}(k) \left[\bar{R}(k+1) + P(k+1) \bar{A}(k) \right] \right\}^{-1} \bar{B}(k) \left[\bar{R}(k+1) + P(k+1) \bar{A}(k) \right] \bar{A}(k) ; \quad (7)$$

the matrix, \bar{P} , satisfies the following recursive matrix Riccati equation,

$$\bar{P}(k) = \bar{A}(k) \left[\bar{R}_1(k+1) + \bar{P}(k+1) \right] \left[\bar{A}(k) - \bar{B}(k) \bar{F}(k) \right] ; \quad (8)$$

the one-step predictor output, $\hat{\bar{X}}$, is defined by

$$\hat{\bar{X}}(k+1) = \bar{A}(k) \hat{\bar{X}}(k) + \bar{B}(k) U(k) + \bar{K}(k) [Y(k) - \bar{C}(k) \hat{\bar{X}}(k)] ; \quad (9)$$

the estimator gain, \bar{K} , satisfies the following equation,

$$\bar{K}(k) = \bar{A}(k) \bar{Q}(k) \bar{C}(k) \left[\bar{C}(k) \bar{Q}(k) \bar{C}(k) + V_2(k) \right]^{-1} ; \quad (10)$$

and, the state estimation error covariance matrix, Q , satisfies the following recursive matrix Riccati equation,

$$\bar{Q}(k+1) = \left[\bar{A}(k) - \bar{F}(k) \bar{C}(k) \right] \bar{Q}(k) \bar{A}(k) + \bar{V}_1(k) . \quad (11)$$

The final value of \bar{P} used to "initialize" Equation (8) (which is solved backward in time) is the final value defined in the quadratic cost function of Equation (5), i.e., $P(k_1) = P_1$. The initial value of Q used to initialize Equation (11) is the error covariance of the initial estimate of X , i.e., $Q(0) = Q_0$.

If the system statistics defined in Equations (1) and (2) are gaussian, then the above solution is the optimal solution without qualification; if not, then it is the optimal linear control solution. The expected system performance is determined by analyzing the augmented system as a linear, stochastic, regulator problem.

It is noted that the analogous, explicit solution to the general continuous-time problem does not exist [2]. However, the analogous continuous-time solution does exist for the special case where only measurement delays are considered [3].

4. PRACTICAL CONSIDERATIONS

From the control gain, Equations (7) and (8), and the estimator gain, Equations (10) and (11), it can be shown that the required dimensions of the controller and estimator are not, in general, equal. The dimensions in the control matrix Riccati equation, (8), is determined by the number of delayed states in the process evolution, i.e., \bar{P} has the dimensions of $n \cdot q \times n \cdot q$. As a result, for the special case of measurement delays only ($q = 1$), the controller implementation is unaffected. On the other hand, the dimensions in the estimator matrix Riccati equation, (11), are determined by the maximum of p and q . If $h = \max(p, q)$, then Q has the dimensions $n \cdot h \times n \cdot h$. Consequently, any system time-delays will increase the dimensions of the estimator; however, only time-delays in the process evolution affect the design of the deterministic controller. For the general case, then, the optimal control is given by:

$$U(k) = - \sum_{i=1}^q F_i(k) \hat{\bar{X}}(k+1-i) \quad (12)$$

where

$$\bar{F}(k) \triangleq \left[F_q(k) \mid F_{q-1}(k) \mid \cdots \mid F_2(k) \mid F_1(k) \right] \quad (13)$$

and

$$\hat{\bar{X}}(k) \triangleq \left[\hat{\bar{X}}(k+1-h) \mid \cdots \mid \hat{\bar{X}}(k+1-q) \mid \cdots \mid \hat{\bar{X}}(k-1) \mid \hat{\bar{X}}(k) \right]^T . \quad (14)$$

We note that, in Equation (14), $\hat{\bar{X}}(k)$ is the one step predicted value of the original system state vector; $\hat{\bar{X}}(k-1)$ is the filtered value; $\hat{\bar{X}}(k-2)$ is the one-step smoothed value; and, finally, $\hat{\bar{X}}(k+1-h)$ is the $(h-2)$ th smoothed value.

Typically, the real time computational requirements associated with the implementation of time varying optimal control is always stressing due to the "backward-in-time" recursion which is required to obtain solutions, P , of Equation (8). However, in many high-accuracy SDI related applications, the A , B , R_1 , and R_2 matrices of the augmented system and cost function can be treated as time invariant over the time intervals of interest. Further, if the augmented system satisfies the relatively minor requirements of stabilizability and detectability, then the control gain, \bar{F} , will converge to a unique value such that the steady-state optimal control law is time invariant, asymptotically stable, and minimizes the quadratic cost function of Equation (5) as $k_1 \rightarrow \infty$. For this case the steady-state gain matrix, \bar{F}_{ss} , can be computed off-line and stored for real-time use; the real-time computations required to implement this steady-state control law are negligible.

Usually, for tracking and regulator problems, the steady-state control gains and time varying gains are such that the initial value of the time varying gain is equal to the steady-state gain but is less than, or equal to, the steady-state gain as time progresses, i.e.,

$$\bar{F}_{ss} \geq \bar{F}(k) \quad \text{for } k_0 \leq k \leq k_1.$$

As a consequence, the steady-state gain tends to maintain the controlled state closer to the estimated state, but at the cost of control energy. If accuracy is the significant criterion, then the steady-state gain is not only easy to implement, it also provides essentially equivalent or better accuracy.

This, however, is not the case for the estimator gain even though, in most cases where linear steady-state optimal stochastic control is implemented, both the steady-state control and estimator gains are used. For the estimator, and the same general conditions as above, the initial time varying estimator gain is usually significantly larger than the steady-state gain to account for initial uncertainties in the knowledge of the system state. As time progresses, the time varying gain converges to the steady-state gain, i.e.,

$$\bar{K}_{ss} \leq \bar{K}(k) \quad \text{for } k_0 \leq k \leq k_1.$$

Consequently, initial system performance will be significantly degraded if \bar{K}_{ss} is used. In fact, if applied to a "linearized" system, the estimator may actually diverge given the initially small steady-state filter gains.

Hence, a good compromise between performance and computational complexity is to choose the steady-state controller with the time varying estimator. Further, since the estimator matrix Riccati equation is solved forward in time, the computations associated with the time varying filter gain are orders of magnitude less than the time varying control gain and can usually be implemented in real time.

While the computations associated with the augmented system estimator are significantly increased because of the increased dimensions, some simplifications can be made. The estimator Riccati equation can be separated into a time update and a measurement update where the time update for the augmented system becomes primarily one of data transfer. If \bar{Q} , from Equation (11), is defined as:

$$\bar{Q}(k) \triangleq \begin{bmatrix} Q_{1,1}(k) & Q_{1,2}(k) & \dots & Q_{1,h}(k) \\ Q_{2,1}(k) & Q_{2,2}(k) & & \vdots \\ \vdots & \ddots & \ddots & \vdots \\ Q_{h,1}(k) & \dots & \dots & Q_{h,h}(k) \end{bmatrix} \quad (15)$$

then, for the special case where $q = 1$ and $p \geq 2$, the time update becomes,

$$\left. \begin{aligned} Q_{i,j}(k+1) &= Q_{i+1,j+1}(k) \\ Q_{i,p}(k+1) &= Q_{i+1,p}(k) A_1^T(k) \\ Q_{p,p}(k+1) &= A_1(k) Q_{p,p}(k) A_1^T(k) + v_1(k) \end{aligned} \right\} \quad i, j=1, \dots, p-1 \quad (16)$$

For the worst case, where $q > p \geq 1$, the time update is given by:

$$Q_{i,j}(k+1) = Q_{i+1,j+1}(k) \quad ; \quad i, j = 1, \dots, q-1$$

$$\left. \begin{aligned} Q_{i,q}(k+1) &= \sum_{j=1}^q Q_{i,j}(k) A_{(q+1)-j}^T(k) \quad ; \quad i=1, \dots, q-1 \\ Q_{q,q}(k+1) &= \sum_{i=1}^q \sum_{j=1}^q A_{(q+1)-i}(k) Q_{i,j}(k) A_{(q+1)-j}^T(k) + V_1(k) \end{aligned} \right\} \quad (17)$$

In practice, another computational simplification results since rarely, if ever, will the actual system measurement be a function of every time-delayed state element and, likewise, for the actual system process model. Hence, the submatrices A_i and C_i (for $i > 1$) in A and C , respectively, will usually be of significantly reduced dimension. This, in turn, significantly reduces the dimensions of the augmented system model.

Finally, we note that the iteration interval, Δt , (implied in the discrete-time system representation) or some integer number of Δt 's, should be set equal to the time-delay. For variable time delays, it may be advantageous to use: a) variable iteration intervals for the estimator, b) fixed iteration intervals for the controller and, c) variable-time updates of the state estimate to time synchronize the estimator output with the controller input.

5. SUMMARY

The optimal control problem has been addressed in which both the measured system output and process evolution are functions of time-delayed states. The optimal discrete-time solution is developed by re-formulating the original system, with explicit time-delayed states, into standard regulator form using state vector augmentation, and then applying the theorems of optimal control. This solution is directly applicable to the tracking problem as well since tracking problems can be re-formulated as regulator problems. Considerations are given for the implementation of this control law, including the justification for a steady-state deterministic controller with state input provided by a time-varying one-step Kalman predictor.

This solution should prove useful in current and future high accuracy guidance and control applications. The analogous solution, in continuous time, does not exist.

REFERENCES

- [1] H. Kwakernaak and R. Sivan, Linear Optimal Control Systems, New York: Wiley-Interscience, 1972.
- [2] H. Kwakernaak, "Optimal Filtering in Linear Systems with Time Delays," IEEE Transactions on Automatic Control, Vol. AG-12, No. 2, April 1967, pp. 169-173.
- [3] D. L. Kleinman, "Optimal Control of Linear Systems with Time-Delay and Observation Noise," IEEE Transactions on Automatic Control, October 1969, pp 524-526.

OVERVIEW OF OMEGA SIGNAL COVERAGE

by

Radha R. Gupta

The Analytic Sciences Corporation
55 Walkers Brook Drive
Reading, Massachusetts 01867
United States

and

Peter B. Morris

US Coast Guard,
Omega Navigation System Center,
7323 Telegraph Road
Alexandria, Virginia 22310
United States

ABSTRACT

Successful use of the Omega Navigation System requires external information on the system coverage (i.e., spatial and temporal accessibility of "usable" signals from the Omega system) in selecting Omega signals for position fix computations. This paper reviews the published Omega signal coverage information and describes in detail the type (and basis) of the coverage information currently disseminated by the U.S. Coast Guard Omega Navigation System Center. Assessments of the worldwide coverage provided by the system, and the impact of Omega transmitting station outages on the system coverage are also given.

1 INTRODUCTION**1.1 BACKGROUND**

The Omega Navigation System is a VLF radionavigation system transmitting navigation signals at 10.2, 11.05, 11.33 and 13.6 kHz from the eight transmitting stations shown in Fig 1-1. The system is designed to provide a worldwide position fix capability, by employing phase measurements from three* or more stations, to an accuracy of 4 nautical miles (95% of the time) under all weather conditions. Knowledge of "system coverage" (i.e., spatial and temporal accessibility of "usable" signals from the Omega system stations) is critical to mission planning, system usage, and system reliability assessment.

The Omega Navigation System Center† (ONSCN), a U.S. Coast Guard Agency, has the primary responsibility for operating, maintaining, improving, and encouraging

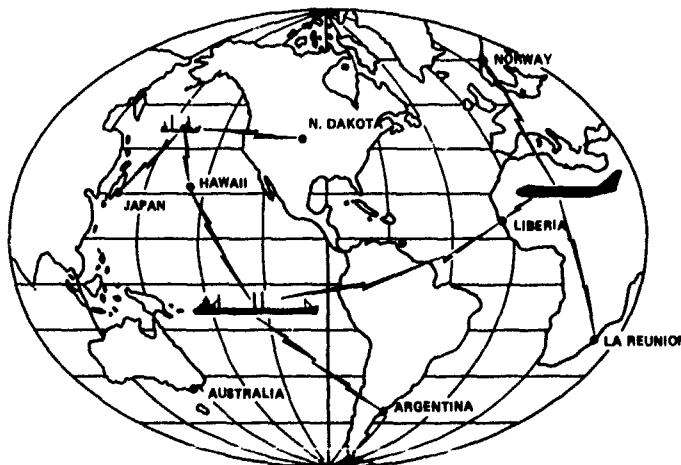


Figure 1-1 Omega Transmitting Stations

*Only two stations are required if the user has a stable (at least one part in 10 million) frequency standard, or if a filtered estimate of the Omega system epoch is available using earlier data from three or more stations.

†Formerly known as the Omega Navigation System Operations Detail (ONSOD).

the use of the Omega system. To aid Omega users in selection of usable Omega signals (see Section 1.2) for their mission, ONSCEN provides the usable signal coverage information. ONSCEN has also undertaken a Regional Validation Program (Ref. 1) to collect Omega coverage data in various navigationally-significant regions of the world. This validation program began in year 1977 and is expected to be concluded in 1990. The validation data along with the data collected at the worldwide-distributed, fixed Omega signal monitoring sites (Ref. 2) are used to validate/refine the predicted coverage boundaries, and to improve the Omega signal propagation prediction models.

As the Omega system has matured with time, receivers have become more sophisticated due to (1) the introduction of inexpensive microprocessors, and (2) the development of improved coverage prediction models. With the help of information derived from the ongoing validation program, coverage information has been refined, augmented, and repackaged to provide better accuracy and improved display.

This paper presents an overview of the available Omega signal coverage information, with emphasis on the coverage information currently disseminated by ONSCEN, and provides an assessment of the worldwide capabilities of the Omega system. Section 1.2 defines usable signals, while Section 2 presents a review of available coverage information. Assessments of world-wide coverage provided by the Omega system and system availability (due to station outages) are given in Sections 3 and 4, respectively. The paper is summarized, along with conclusions, in Section 5. An overview of the Omega signal propagation mechanism is given in Appendix A. A discussion of the coverage prediction models used for developing the signal coverage information is provided in Appendix B.

1.2 USABLE SIGNALS

An Omega station signal received at a specific (geographic) location and time is usable if the signal.

- Is sufficiently strong such that the signal-to-noise ratio (SNR) is greater than the receiver detection threshold
- "Non-modal", i.e., the signal phase vs distance relationship does not deviate significantly from the "Mode 1" signal phase relationship.

The phase deviation is due to presence of strong-amplitude higher-order (i.e., non-Mode 1) modes in a signal, and is commonly referred to as "modal interference" (MI)-induced phase deviation, or simply MI. In addition, the signal must also be a short-path[†] signal, as most Omega receivers employ a short-path phase prediction algorithm.

2 AVAILABLE SIGNAL COVERAGE INFORMATION

A historical time-line of the various types of Omega coverage information published by various researchers is given in Table 2.1. Bortz et al. (Ref. 3), in 1976, were the first to provide coverage information in the form of "individual station" and "composite" diagrams. An individual station diagram displays predicted worldwide coverage of usable signals from a station at a specified time for a given usable signal criteria; a composite diagram displays the same information for the full network of Omega system stations. These diagrams were developed for the 10.2 kHz signal coverage for two extreme local temporal conditions: local summer noon and local winter midnight. In these diagrams, the usable signal has: (1) SNR \geq -20 dB (in a 100 Hz noise bandwidth), and (2) phase deviation, $\Delta\phi$, less than 13 centicycles (cec). These diagrams were not validated as the required empirical data were not readily available at that time.

*See Appendix A.

[†]A short path is the shorter of the two great-circle arcs over the surface of the earth joining the transmitting station and receiver

TABLE 2-1
TIME-LINE OF PUBLISHED COVERAGE INFORMATION

YEAR DEVELOPED/ PUBLISHED	COVERAGE ATTRIBUTES					RESEARCHERS
	FREQUENCY (kHz)	OMEGA STATION POWER*	TIME	EMPIRICAL VALIDATION	DISPLAY MEDIUM	
1976	10.2	Full	Two Local Times of Day	None	Hardcopy	Bortz <i>et al</i> (Ref. 3)
1980	"	"	Eight Global Times of Year	Some	"	Gupta <i>et al</i> (Refs. 4 through 10)
1983	"	"	Minimum and Maximum Coverage Contours	"	Algorithm	Gupta and Warren (Refs. 11)
1983	"	"	7-vo Local Times	"	Hardcopy	Swanson (Refs. 12 and 13)
1985	"	"	Several Global Times of Year	"	"	Swanson (Refs. 14 and 15)
1985	13.6	"	Eight Global Times of Year	Substantial	"	Gupta <i>et al</i> (Refs. 16 through 18)
1986	10.2 & 13.6	Full & Reduced	"	"	Micro- computer Display	Warren <i>et al</i> (Ref. 19)

*Full power is 10 kW radiated power

Further, these diagrams have had limited application because of the overly conservative $\Delta\phi$ criterion and unrealistic illumination conditions employed

In 1978, advent of inexpensive microprocessors led to use of multiple-frequency signals in position-fix computations, yielding higher accuracy fixes than were possible with the use of single-frequency signals. This development promoted use of the Omega receiver as an inexpensive alternative to the inertial navigation system for airborne enroute navigation. As a consequence, more airborne users began using Omega and thus needed signal coverage information for all hours of the day. Although it would have been highly desirable to have coverage information for each hour of the day, the cost of development and dissemination of such information in a comprehensive way was deemed prohibitive at that time. In response to Omega user needs, ONSCEN (then known as ONSOD) sponsored development of global time-specific coverage information, for eight representative global times of the year, for 10.2 kHz (in 1979) and 13.6 kHz (in 1983) by Gupta *et al* (Refs. 4 through 10, and 16 through 18). This information was developed for the more realistic phase deviation criterion of 20 cec (as opposed to the 13 cec criterion used by Bortz *et al* (Ref. 3)). This coverage information is currently disseminated by ONSCEN to aid Omega users in selection of the usable 10.2 and 13.6 kHz signals.

In the Gupta *et al* diagrams (individual station and composite), a station signal at a location/time is considered usable if the received signal simultaneously satisfies both the SNR and $\Delta\phi$ criteria given in Table 2-2 at each point along a minimum length radial path segment through the point. The minimum length was chosen to be one megameter (1000 km) in the 10.2 kHz diagrams, and two megameters in the 13.6 kHz diagrams. In addition, a station's signals are considered unusable near the station antipode[†], because the path azimuth changes too rapidly near the antipode and thus one may quickly pass from short-path to long-path conditions; therefore, a region of one megameter surrounding the station

*The 13 cec $\Delta\phi$ criterion was selected because this corresponded to a hyperbolic LOP error of one nautical mile (at 10.2 kHz) on the baseline (assuming one of the LOP station has no phase error). In the later diagrams, a less conservative criterion of 20 cec is used which correspond to just under 1/4 cycle (25 cec) phase error which, with the aid of a phasor diagram, can be shown to be the threshold at which cycle jumps occur.

†A station antipode is the point on the earth's surface diametrically opposite the station location.

TABLE 2-2
ATTRIBUTES OF GUPTA et al. COVERAGE DIAGRAMS

COVERAGE ATTRIBUTES						DIAGRAMS PUBLICATION REFERENCES†	
SIGNAL FREQUENCY (kHz)	COVERAGE DIAGRAM TYPE	USABLE SIGNAL CRITERIA*		TIMES	DIAGRAM PROJECTION	JOURNAL	REPORT
		SNR** ≥	$\Delta\phi \leq$				
10.2	STATION SIGNAL COVERAGE	TWO CRITERIA	20 cec	0600 & 1800 GMT IN FEB. MAY AUG. & NOV.	MERCATOR	4,5	6,7,8,9
	COMPOSITE SIGNAL COVERAGE	-20 dB & +30 dB			MERCATOR & POLAR		
	STATION NIGHTTIME MODAL INTERFERENCE	X	20 cec	WORLDWIDE NIGHTTIME CONDITIONS	MERCATOR	10	6
13.6	STATION SIGNAL COVERAGE	-20 dB	20 cec	0600 & 1800 GMT IN FEB. MAY AUG. & NOV.	MERCATOR	16,17	18
	COMPOSITE SIGNAL COVERAGE				MERCATOR & POLAR		
	STATION NIGHTTIME MODAL INTERFERENCE	X	20 cec	WORLDWIDE NIGHTTIME CONDITIONS	MERCATOR	10	6

*Signal must be Mode 1-dominated signal

†See references at the end of paper

**SNR in 100 Hz noise bandwidth

antipode is excluded from the station usable signal coverage considerations. Because Omega signals are severely modal (i.e., $\Delta\phi > 20$ cec) in a station "near-field" region (which was assumed to extend to a minimum distance of 500 km from the station at 10.2 kHz, and 1000 km at 13.6 kHz), $\Delta\phi$ was not computed but was assumed to be greater than 20 cec in the station near-field region. Samples of the individual station and composite diagrams are shown in Figs. 2-1 and 2-2, respectively. The eight coverage times (see Table 2-2) are selected to maximize

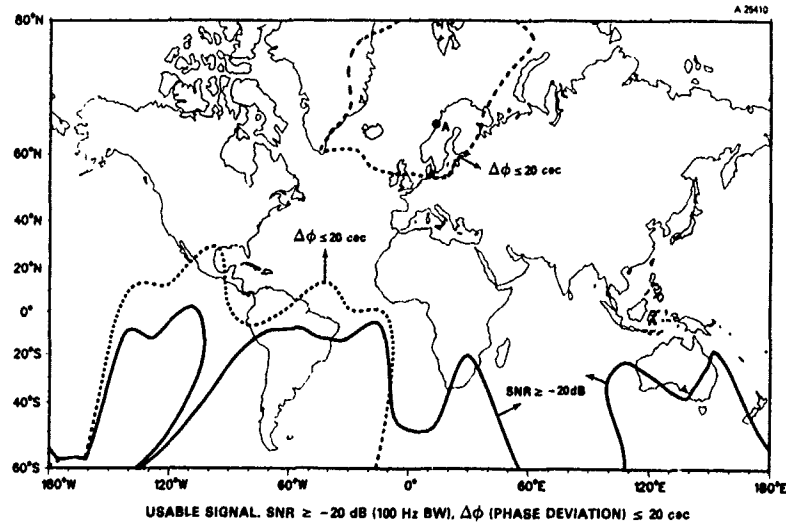


Figure 2-1 An Example of an Individual Station Coverage diagram: Norway Station, 13.6 kHz, February 06000 GMT (Ref. 18)

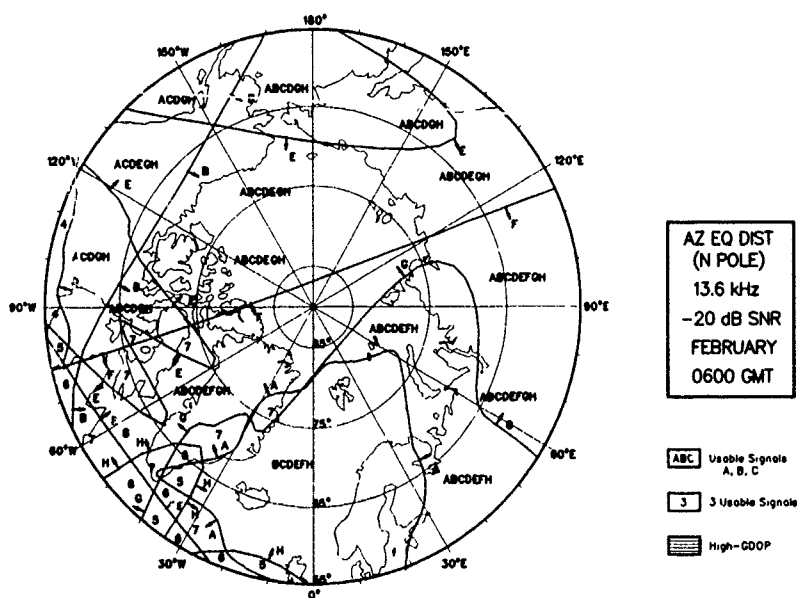
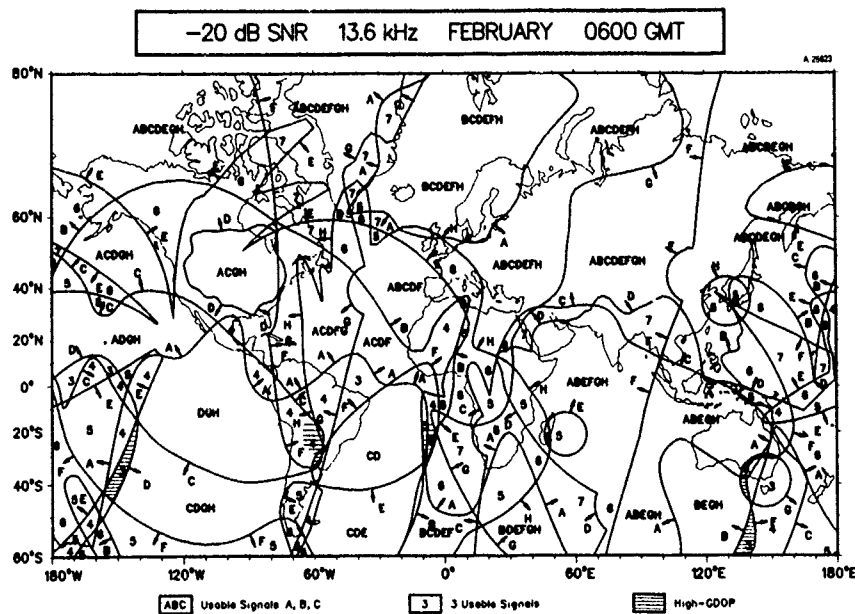
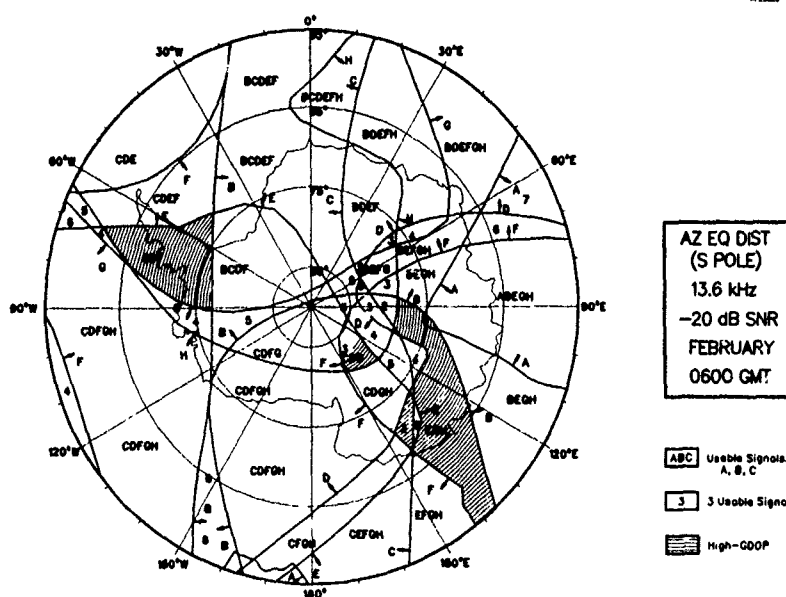


Figure 2-2 Examples of Composite Signal Coverage Prediction Diagrams:
13.6 kHz Coverage at 0600 GMT in February (Ref. 18)

the shortest distance of the day/night terminator from all eight Omega stations at all eight coverage times; the shortest station-to-terminator distance is 700 km.

In addition to displaying the Omega system signal coverage, the composite diagrams (developed by Gupta *et al.*) identify regions of the world where three or more usable signals are simultaneously available at a given time, but none of these signals at that time provide a 3-station hyperbolic fix with a geometric-



(c) South Pole-Centered Azimuthal-Equal-Distance (AED) Projection

Figure 2-2 Examples of Composite Signal Coverage Prediction Diagrams.
13.6 kHz Coverage at 0600 GMT in February (Ref. 18) (Continued)

dilution-of-precision (GDOP) less than a prescribed threshold value. A threshold GDOP value of one kilometer of radial position error per centicycle of phase-difference error (standard deviation) is used in the 10.2 kHz diagrams, and a value one-half as large is used in the 13.6 kHz diagrams. Although the composite diagrams serve as useful purpose, i.e., display simultaneous availability of usable signals from the Omega system stations, the presentation format of these diagrams makes their use somewhat difficult, especially for the first-time users. A more appealing and useful presentation format/medium is thus highly desirable. Progress in presentation techniques have recently been made by Swanson (Ref. 15) and by Warren *et al.* (Ref. 19), as discussed later in this section.

The individual station nighttime modal interference prediction diagrams, developed by Gupta *et al.* (Refs. 8, 10 and 18), identify geographic regions of the world where the station signals at a given frequency are predicted to be "non-modal" (i.e., phase deviation $\Delta\phi \leq 20$ cec) for assumed worldwide nighttime illumination conditions. Since modal effects along a signal path generally vary with changing illumination condition along the path and are (usually) most severe when the path has "all-night" illumination, any region identified to be non-modal for the nighttime condition is likely to remain non-modal for most other path illumination conditions. The nighttime modal interference diagrams provide useful guidance for deselection of modally-disturbed signals. Individual station nighttime modal interference diagrams for 10.2 and 13.6 kHz signals are available from ONSCEN. An example of the combined 10.2 and 13.6 kHz frequency nighttime modal interference diagram for the Liberia station is shown in Fig. 2-3.

The signal coverage and modal interference diagrams developed by Gupta *et al.* have been validated with coverage data collected under the ongoing Omega Regional Validation Program (Ref. 1). In addition, these diagrams have been validated with data collected by the worldwide network of fixed Omega monitor sites maintained by ONSCEN (Ref. 2).

The coverage diagrams, though not fully validated* with empirical data, have proven, for the most part, to be quite reliable. They provide useful and reliable guidance to worldwide Omega users in selection of usable signals, thereby permitting more accurate navigation and positioning information. The utility of these diagrams could be enhanced by considering: (1) the impact of long-path signals on the usable signals in the diagrams, and (2) the worldwide accuracy that the Omega system is capable of providing through optimal processing of all usable signals available from the Omega system stations at all Omega frequencies.

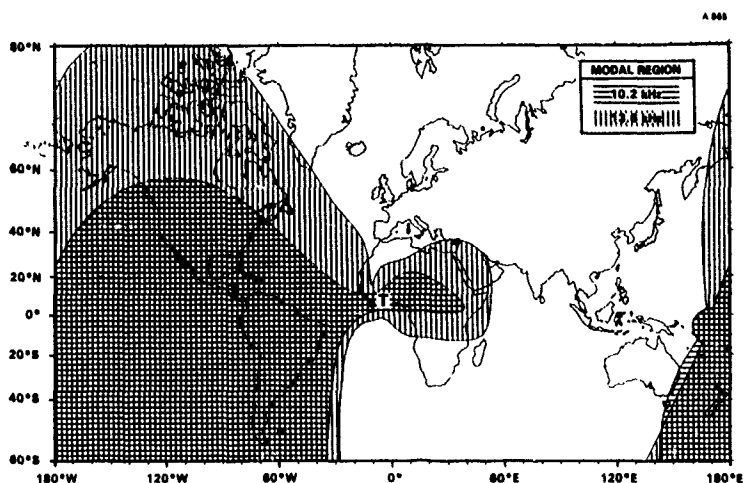


Figure 2-3 An Example of a Combined 10.2 and 13.8 kHz Frequency Nighttime Modal Interference Diagram: Liberia Station (Ref. 10)

In view of the general availability of the microprocessor-based Omega receivers and the potential for incorporating the Omega coverage information in these receivers to select usable signals, Gupta and Warren (Ref. 11), in 1983, published a simple algorithm for incorporating coverage information in such receivers. The algorithm provides minimum and maximum coverage extremes of usable 10.2 kHz signals from individual Omega stations.

The coverage extremes of a station, over the set of eight global times/months considered here, (Fig. 2-4) are determined by overlaying the eight global-time-specific coverage contours (from Gupta et al. diagrams, Ref. 7) of the station and then finding the inner and outer coverage extremes. The algorithm provides range and bearing coordinates of the inner and outer coverage boundaries. Because of the very limited data storage requirements of the algorithm, it is particularly attractive for use in microprocessor-based Omega receivers.

In 1983, Swanson (Refs. 12 and 13) published a set of 10.2 kHz signal coverage and accuracy diagrams for individual Omega stations and the full Omega system for "idealized-day" and night conditions, using a "parametric approach". This approach uses a two-mode (Modes 1 and 2) parametric model for signal amplitude and phase deviation predictions, and has since been extended to provide 10.2 kHz coverage predictions at fixed global times (Refs. 14 and 15). This method is also being extended to provide 13.8 kHz signal coverage predictions. The parametric model does not fully characterize the nighttime, transequatorial propagation effects along westerly paths, as predicted by the full-wave IPP model (Ref. 20). Furthermore the signal phase deviation predictions ignore mode conversion effects occurring at the path-(day/night) terminator discontinuities along mixed paths, which are included in the Gupta et al. coverage information for 13.8 kHz. Individual station signal coverage information is provided at worldwide (10 deg by 10 deg latitude-longitude) grid points (Fig. 2-5) where the notation on each grid point indicates limitations (if any) on the station signal due to signal-to-noise ratio, modal interference, long-path interference, and atipodal effects. The individual station coverage information for the full system has been combined with the expected random errors in the signal phase measurements to develop Omega system accuracy diagrams. The position fix accuracy available from the full system through "optimum" use of available station signals is indicated at the worldwide latitude-longitude grid points. An example of a system accuracy diagram of this type is shown in Fig. 2-6.

*Especially the Australia station signal coverage, as this station was the last station to join the system, becoming operational in August 1982.

†The usable signals are the short-path signals satisfying specified usable signal access criteria.

‡Bias errors are ignored.

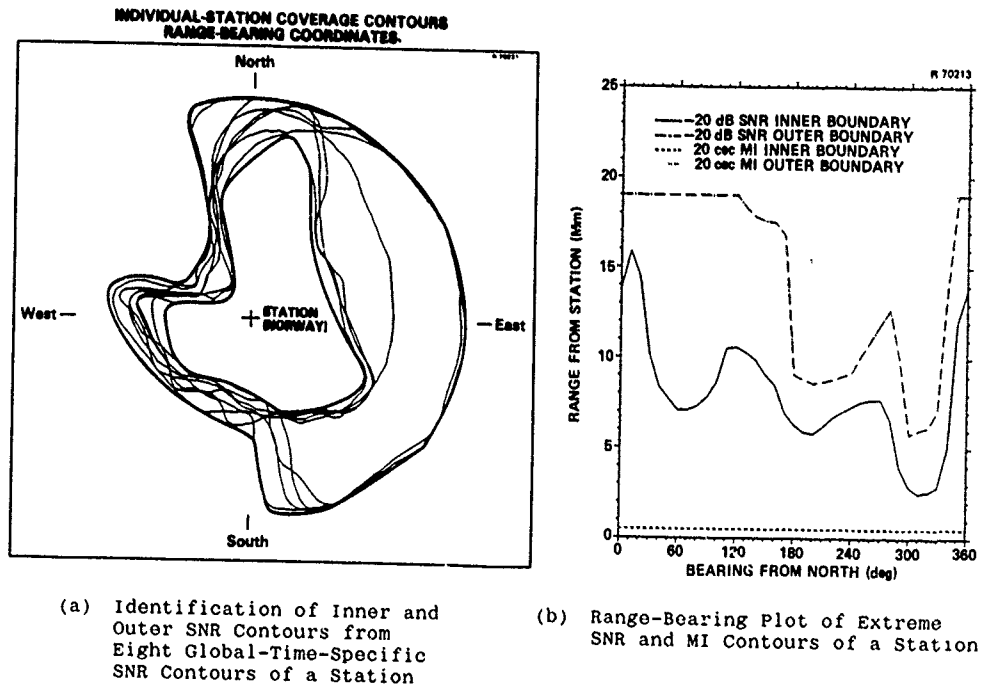


Figure 2-4 An Example of Extreme SNR and MI Contours of a Station (Ref. 11)

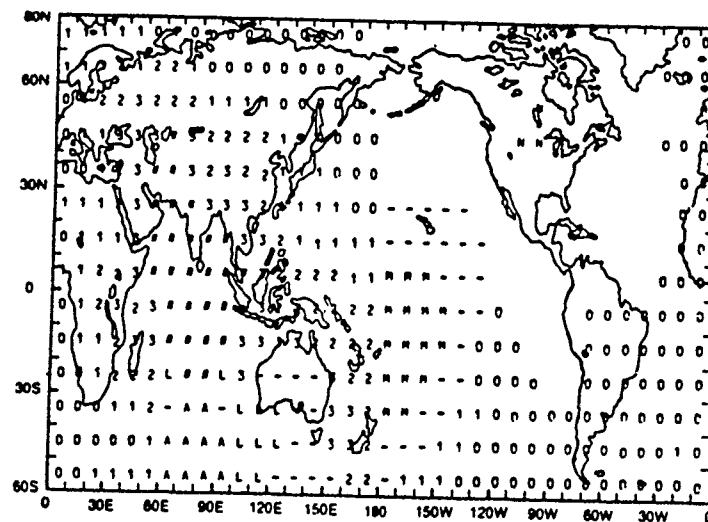


Figure 2-5 An Example of Swanson's (Parametric Approach-Based) Individual Station Coverage Diagrams: North Dakota at 0600 GMT (Vernal Equinox) (Key: Self-Interference; M = Modal, L = Long-Path, N = Station Near-Field, A = Antipodal, - = Perturbed but Usable; SNR: (Blank) = $\text{SNR} \geq 0$ dB (in 100 Hz Bandwidth), 1 = $-10 \leq \text{SNR} < -20$ dB, etc., # = $\text{SNR} \leq -40$ dB) (Ref. 16)

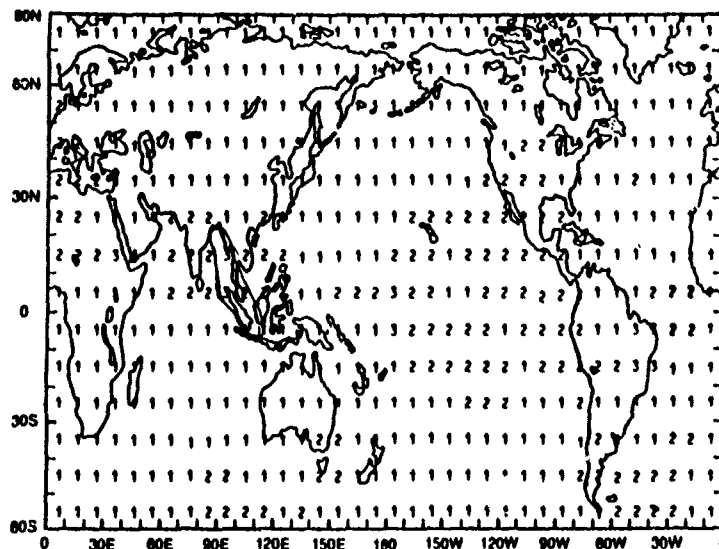


Figure 2-6 An Example of a Swanson (Parametric Approach-Based) Coverage Diagram for 24-hour System Accuracy at 10.2 kHz (Numbers Show Median Accuracy in nautical miles, cep) (Ref. 16)

Swanson's individual station coverage diagrams add a dimension of information not included in the diagrams published by Gupta *et al.*, as Swanson has incorporated information on the presence of (undesired) strong-amplitude long-path signals that may interfere with, or dominate, the (desired) short-path signals. Although a number of individual station coverage and Omega system accuracy diagrams have been published by Swanson (Refs. 12 through 15) for 10.2 kHz signals, a full set of validated diagrams is not available for worldwide coverage/accuracy guidance at 10.2 kHz for all global times. The 13.6 kHz diagrams are currently being developed by Swanson (Ref. 15).

The spatial and temporal complexities of system coverage make it difficult to rapidly assimilate and make effective use of the extensive information contained on the hardcopy diagrams for side-by-side comparisons, or to perform "what-if" analyses of non-standard conditions so as to assess mission planning alternatives. This has led to development of a microcomputer display medium for signal coverage information. This software tool, developed by Warren *et al.*, (Ref. 19) is called Omega Automated Composite Coverage Evaluator of System Signals (Omega ACCESS). An overview of Omega ACCESS is given in Fig. 2-7. Omega ACCESS displays coverage data in several world map projections (Mercator, polar, station-centered Azimuthal-Equal-Distance (AED), station-antipode-centered AED, etc.), and allows the user/analyst to answer a wide variety of "what-if" coverage questions related to changing propagation scenario/conditions. A significant feature of Omega ACCESS is the easy update of coverage information to incorporate new findings from ONSCEN's ongoing Omega Regional Validation Program, and to reflect any future changes in the system configuration/operation. This will allow faster turnaround and distribution of the latest coverage information to the user community.

The current coverage data base incorporated in Omega ACCESS is the same as was used by Gupta *et al.* to develop the 10.2 and 13.6 kHz signal coverage diagrams (Refs. 4 through 10, and 16 through 18) except for some minor adjustments. These adjustments were applied to 10.2 kHz signal coverage contours in the South Atlantic region to reflect recent validation program results in the region. In addition to nominal (10 kW) station radiated power coverage data, Omega ACCESS provides coverage information for reduced station power conditions. Specifically, coverage diagrams are provided for: (1) a 6 dB reduction (from the nominal) in the radiated power level of each Omega station excepting the Japan station, and (2) 2, 4, 6, 8 and 10 dB reductions in the Japan station radiated power level.

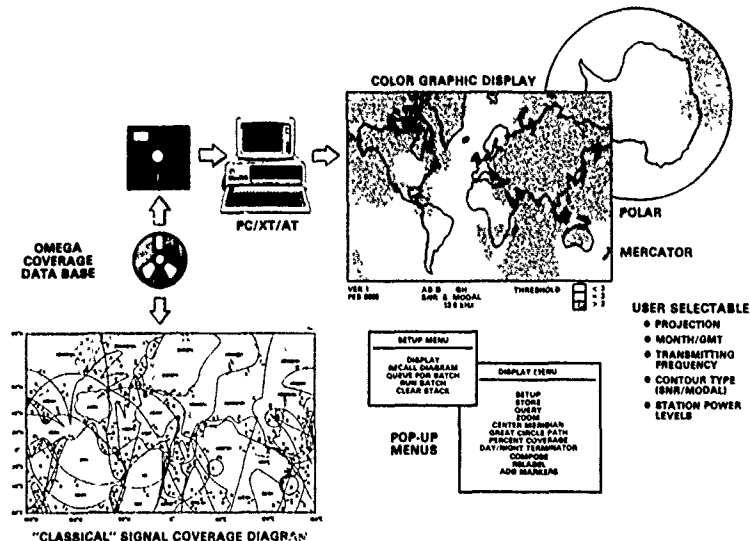


Figure 2-7 Overview of Omega ACCESS (Ref. 19)

3. OMEGA SIGNAL COVERAGE ASSESSMENT

This section presents general characteristics of individual station signal coverage, and provides an assessment of composite (full system) coverage, based on the Gupta *et al.* coverage information.

3.1 INDIVIDUAL STATION SIGNAL COVERAGE CHARACTERISTICS

Examination of the individual station signal coverage information reveals that:

- Usable signal coverage of a station extends to shorter distances along a day path than along a night path due to the higher signal attenuation rate along the path during day
- Due to geomagnetic field effects, usable signal coverage of a station generally extends to:
 - longer distances in the easterly direction
 - shorter distances in the westerly direction
 - moderate distances in the northerly/southerly direction
- Usable signal coverage does not extend to large distances in Greenland/Antarctica due to extremely high signal attenuation rate (especially under daytime conditions) caused by the very low ground conductivity of these regions
- 13.6 kHz SNR coverage contours extend significantly farther from the signal station than the 10.2 kHz SNR coverage contours, due to a lower attenuation rate at 13.6 kHz
- The largest region(s) with severe phase deviations (i.e., severe modal interference effects) are associated with stations located closest to the geomagnetic equator, such as the Liberia and Argentina stations
- Modal interference at a 13.6 kHz is generally more extensive (spatially) than at 10.2 kHz
- Nighttime modal interference regions extend to the south and west beyond the geomagnetic equator for northern hemisphere stations and to the

north and west beyond the geomagnetic equator for southern hemisphere stations.

3.2 COMPOSITE COVERAGE ASSESSMENT

Based on a study of the composite coverage information, the Omega system coverage, at each Omega navigation frequency, for moderate performance (i.e., -20 dB SNR detection threshold) receivers is almost worldwide; i.e.,

- Over 92% of the earth's surface has usable signals most of the time from at least three stations with good position-fix geometry
- Over 80% of the earth's surface has usable signals most of the time from at least four stations with good position-fix geometry.

For high performance (i.e., -30 dB SNR detection threshold) receivers, the amount of coverage increases to 98% for 3 or more usable signals, and 92% for 4 or more usable signals.

Although the percentage of the earth's surface covered is similar at each frequency, specific geographic regions covered vary. Also, there are several areas of the world where the coverage is either marginal (i.e., coverage exists for only part of the day) or is inadequate. Some of these areas are: the well-known "Winnipeg hole" (in central Canada), and certain regions near the geomagnetic equator and in the general coastal regions surrounding Greenland and Antarctica. These as well as other areas are predicted in the coverage diagrams provided by Gupta et al. (Refs. 4 through 6, 8, 16, and 18).

Also note that although the coverage information is strictly valid for only the stated times (0600 and 1800 GMT in the months of February, May, August, and November), in the lower- and middle-latitude regions of the world the coverage can be more reliably interpolated between the stated months at the same GMT than between the stated GMTs of the same month. This is due to the fact that within these regions, the solar illumination changes much less over three months at a given GMT than over a 12-hr interval in a given month. Coverage data at additional times of the day are required to allow reliable interpolation to other times of the day.

4. OMEGA SYSTEM AVAILABILITY ASSESSMENT

Omega system "availability" (i.e., fraction of the earth's surface over which usable signals are available from the full system multiplied by the fraction of the year during which the stations are on-air) is a function of user navigation mode (hyperbolic or rho-rho), spatial availability (spatial coverage) of the system, and temporal availability (station on-air time statistics) of the Omega transmitting stations. Gupta (Ref. 21) conducted an extensive investigation of the system availability at 10.2 kHz. His findings, based on the available system coverage information at the eight times of the year and station on-air statistics for 1979 and 1980 (Refs. 22 and 23), suggest that the Omega system availability for moderate* performance receivers at 10.2 kHz is about:

- 90% for hyperbolic navigation
- 95% for rho-rho navigation.

Based on the coverage assessments by Gupta et al. (Refs. 4 and 9), the Omega system availability for the high* performance receivers at 10.2 kHz is projected to be about 95% for hyperbolic navigation and about 98% for rho-rho navigation. Since the system coverage does not differ significantly from one Omega frequency to other, similar system availability is expected at all Omega signal frequencies.

5 SUMMARY AND CONCLUSIONS

The Omega Navigation System is the only long-range radionavigation system which provides a continuous position fix capability throughout the world to air-

*Moderate and high performance receivers are assumed to have the SNR detection threshold of -20 dB and -30 dB, respectively, in a 100 Hz noise bandwidth.

borne, marine and terrestrial users, until the U.S. NAVSTAR/Global Positioning System becomes fully operational. The Omega system is expected to support worldwide navigation requirements into the next century.

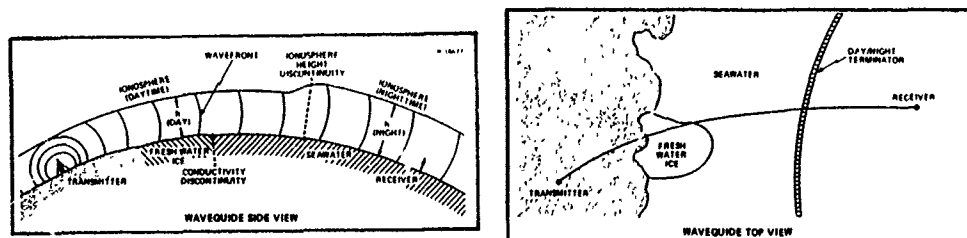
Omega signal coverage is spatially and temporally complex, thereby providing the requirement for signal coverage diagrams to support mission planning, etc. by Omega users. Omega system coverage is a function of the system navigational signal frequency, time (GMT and month), receiver signal-to-noise ratio detection threshold, modal interference rejection threshold, short-path and long-path signal interference rejection threshold, antipodal interference criterion, and geometry between fix-signal stations and user. The system coverage is almost worldwide, although there are several navigationally-important areas of the world (e.g., the Winnipeg hole in central Canada) where Omega coverage is inadequate and assistance from other navigational aids (such as a dead reckoning system) is required to navigate in these areas.

Significant improvements have been made over the past decade in both the Omega receiver technology and Omega coverage accuracy/information-content/presentation. With the advent of inexpensive microprocessors, receivers can now select all of the available usable signals, and compute position fixes using very sophisticated signal phase prediction models. The fixes achieved by these receivers are computed by a "near-optimal" combination of all available usable signals from all Omega stations at all Omega frequencies. Such multiple-frequency fixes have much higher accuracy than that is possible with the single-frequency receivers. Advancement in the area of coverage presentation is evidence in Omega ACCESS (a microcomputer display medium) developed by Warren *et al.* (Ref 19). Omega ACCESS displays coverage information in a number of useful and user-friendly formats, and allows the user/analyst to answer a wide variety of "what-if" questions on coverage as related to changing propagation scenario/conditions.

In the near future, it will be highly desirable to have a validated data base of Omega coverage/accuracy information as a function of time (e.g., for each of 24 GMTs of the day during four representative months of the year) that can be conveniently displayed (and used for analysis) on a microcomputer. The Omega Navigation system Center is currently developing such a coverage information data base and will be disseminating this data base to Omega users.

APPENDIX A OMEGA SIGNAL PROPAGATION MECHANISM

Omega signals propagate in the space between the earth's surface and the upper limit of the ionospheric D-region, referred to as the "earth-ionosphere" waveguide. Signal propagation along a path (see Fig. A-1) is conveniently described as a sum of the "characteristic modes" of the waveguide formed along the path. The locally-varying electromagnetic properties (ground conductivity, solar illumination, geomagnetic field, and direction) of the path determine the propagation characteristics (i.e., amplitude and phase) of the component modes of the signal and hence the resultant signal. Individual mode characteristics are determined by attenuation rate, phase velocity and excitation factor (amplitude and phase) of the mode signal.



(a) Waveguide Side View

(b) Waveguide Top View

Figure A-1 A Path-Waveguide Geometry

Normally, an Omega station signal outside the station "near-field" region* can be approximated by the signal's dominant mode, usually "Mode 1"; that is, the presence of higher-order modes in the signal can be effectively ignored. Mode 1 is the lowest phase-velocity transverse-magnetic mode, and usually has the smallest attenuation rate. The phase of a Mode 1 signal (and hence the phase of a Mode 1-dominated signal) varies almost linearly with distance from a station

In a station near-field region, as well as along nighttime paths at certain azimuths from a station (especially those closest to the geomagnetic equator), Mode 1 often fails to be the dominant signal mode. This lack of Mode 1 dominance in a signal is termed modal interference, and the resulting "modally-disturbed" signal is composed of several competing modes which may alternately dominate on different segments of the path. In this case, due to differing phase velocities of the signal's component modes, the amplitude and phase of the total (multimode) signal exhibit oscillatory behavior with distance. Alternatively, the modally-disturbed signal may be dominated by a single, higher-order mode (e.g., "Mode 2") whose linearly-varying phase vs distance relationship is usually significantly different from that of the Mode 1 relationship (embodied in Omega navigation algorithms).

Modal interference at a path point is characterized as "spatial" or "temporal", depending upon the solar illumination condition along the path between the station and the point. Modal interference during day or night path conditions is classified as spatial interference if the magnitude of the interference at each path point is nearly constant in time (and thus depends only on spatial coordinates). If the path has a mixed illumination condition (i.e., a day/night terminator crosses the path), the interference is referred to as temporal interference. The magnitude of temporal interference at a path point generally varies in response to the movement of the terminator along the path

A.1 SPATIAL MODAL INTERFERENCE

Signal phase deviations arising from spatial interference are generally larger in magnitude and persist to longer distances along a signal path during night than during day. This can be seen by comparing theoretically-predicted behaviors, shown in Figs. A-2(a) and A-2(b),† of 13.6 kHz signals along the northerly-directed, day and night radial paths from the Liberia station. Furthermore, modal effects are usually more severe in magnitude and spatial extent for nighttime paths emanating from transmitting stations located at low geomagnetic latitudes. Figure A-3 shows an example of theoretically-predicted spatial interference effects along a nighttime radial path from the Hawaii station, where "severe" modal effects persist along the entire length of the radial path. In this example, Mode 3 is dominant out to about 3500 km; beyond this point, Mode 2 is the dominant mode of the signal.

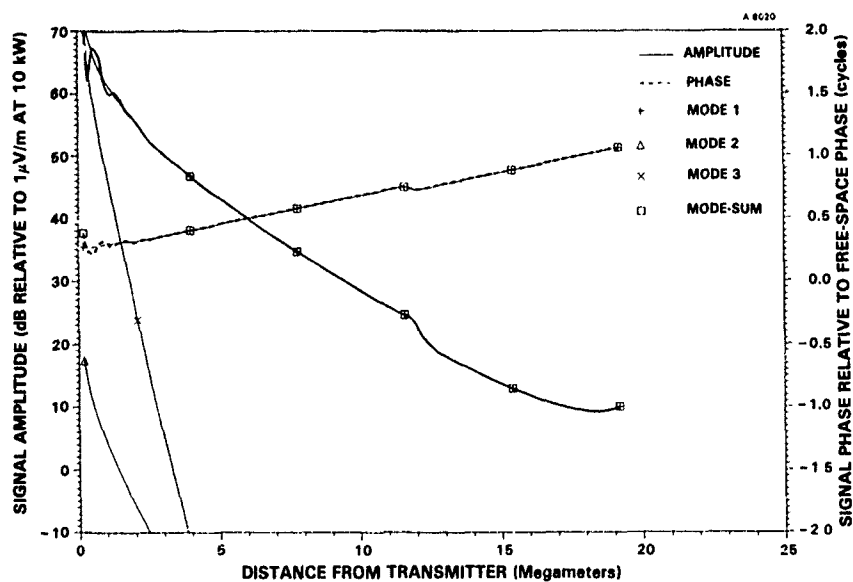
A.2 TEMPORAL MODAL INTERFERENCE

Temporal modal interference arises from "mode conversion" at the day/night terminator crossing (discontinuity) along a mixed path (Fig. A-4) which is subject to the "usual" spatial interference when fully dark. Due to mode conversion, the energy of each incident mode at the path-terminator crossing is converted, or distributed, into several transmitted modes (reflected modes are assumed to be negligible). These transmitted modes propagate beyond the terminator. Thus, because of mode conversion, an Omega station signal with a single dominant mode propagating toward a terminator can exhibit significant modal interference effects (i.e., presence of several competing modes) in the vicinity of, and beyond, the terminator (along the path).

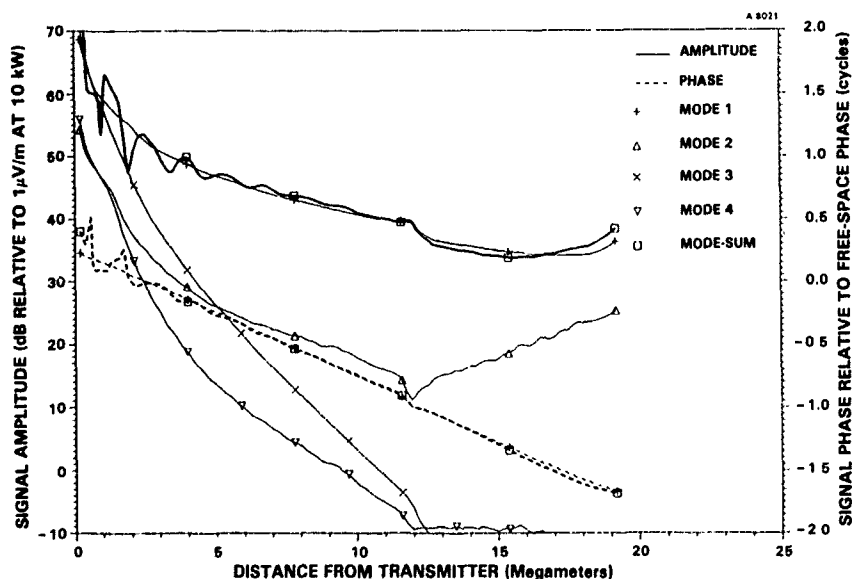
The impact of mode conversion in a propagating signal can be seen in Fig. A-5, which depicts theoretical behavior of 13.6 kHz signal amplitude and phase along the mixed path (at 0800 GMT in May) emanating at 240 deg from the Norway station; the signal behavior was obtained using the theoretical models (Refs. 20 and 24) described in Appendix B. In this illustration, the day/night

*A station near-field region typically extends from the station outward to distances of 500-1000 km along day paths, and 1000-2000 km along night paths.

†The day in Fig. A-2(a) and night signal in Fig. A-2(b) are, respectively, assumed to be composed of the first three (for the day signal) and the first five (for the night signal) lowest phase-velocity modes excited in the earth-ionosphere waveguide modeling and signal path



(a) Day Path



(b) Night Path

Figure A-2 Examples of Theoretical Multimode Signal Propagation along Day and Night Paths (Ref. 20)

terminator along the path is approximately 2.4 megameters from the station, and the signal is assumed to be composed of Modes 1, 2, and 3 along the entire mix path. As expected, the transmitter-excited signal along the day-segment of the mixed path is a predominantly Mode 1 signal. Because of the mode conversion at the path-terminator crossing, the signal beyond 5 megameters from the terminator is no longer a Mode 1-dominated signal. Beyond this distance, the signal is modally-disturbed and is dominated by either Mode 2 or Mode 3.

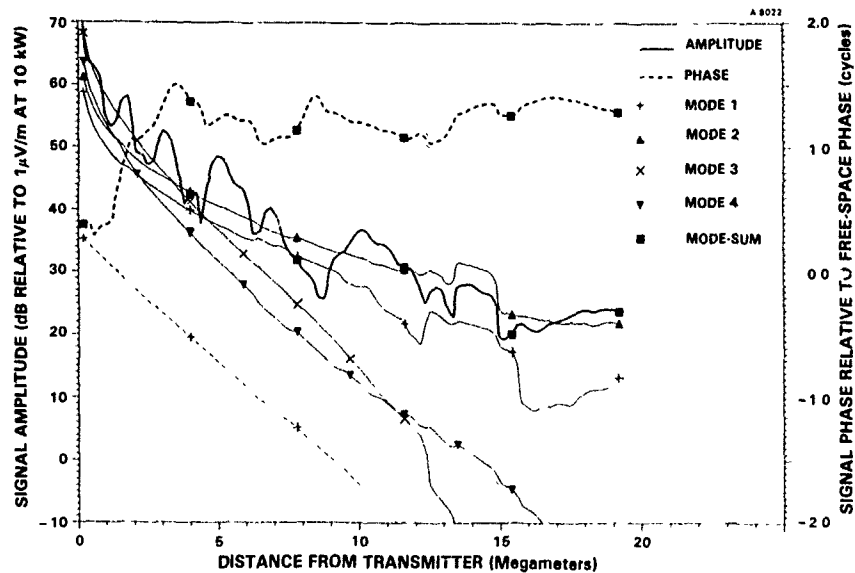


Figure A-3 An Example of Theoretically-Predicted Severe Spatial Modal Interference along a Nighttime Path (Ref 20)

A 5787

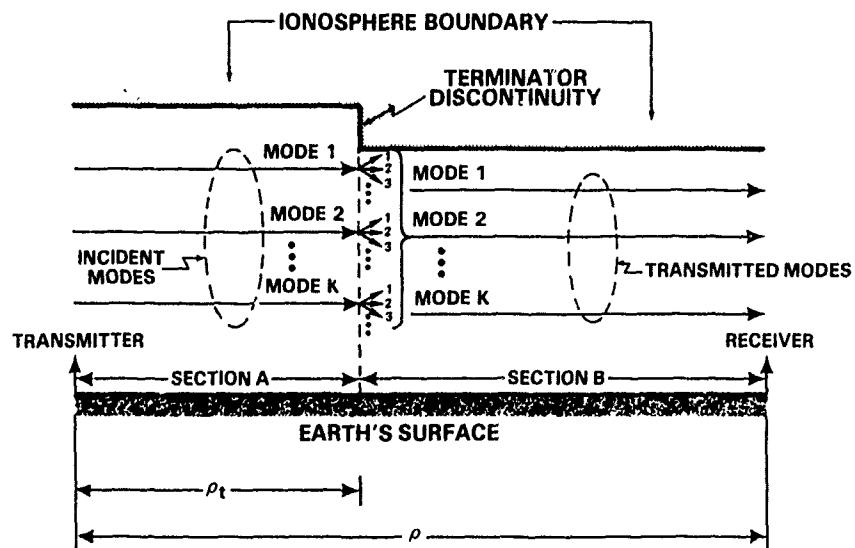


Figure A-4 A Mixed-Path Earth-Ionosphere Waveguide

APPENDIX B OMEGA SIGNAL COVERAGE PREDICTION MODELS

Prediction models are needed to develop Omega signal coverage information. These models must compute amplitude and phase deviation of the Omega signals, and atmospheric noise amplitudes as functions of (geographic) location, frequency, and time. An overview of the prediction models used by Gupta *et al.* (Refs. 4 through 10, and 16 through 18) to generate the global-time-specific coverage information (disseminated by the Omega Navigation System Center) is provided in this appendix.

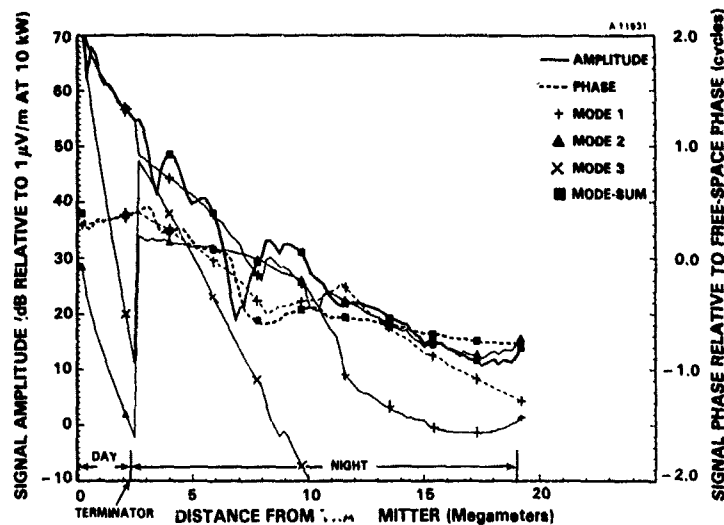


Figure A-5 An Example of Theoretically-Predicted Mode Conversion Effects Along a Mixed Path (Refs. 20 and 24)

B.1 SIGNAL AMPLITUDE AND PHASE DEVIATION PREDICTION MODELS

Of the available very low frequency (VLF) signal prediction models, the mode conversion-theory based model, FASTMC (Ref. 24), is the most accurate for predicting signal behavior along arbitrarily-illuminated paths. This model is an excellent research tool, but not very useful for production work (such as required for developing worldwide coverage information) because of computational cost. The most commonly-exercised VLF signal prediction model is the theoretical IPP model (Ref. 20) which uses a WKB-type path averaging, rather than the mode conversion approach of FASTMC, for determining signal behavior along paths with gradually-varying properties. As a consequence, IPP is applicable to either day or night paths. Thus IPP, supplemented with a diurnal sub-model and a mode conversion model as described below, was used to compute the signal amplitude and signal phase deviation, respectively.

B.1.1 Signal Amplitude Prediction Model

A semi-empirical model was used to compute signal amplitude, which combines the theoretical IPP model-provided Mode 1 spatial sub-models* for day and night paths with an empirically-derived sub-model for transition illumination. The unknown coefficients in the spatial sub-models were determined from the IPP-derived signal amplitude data for day and night paths. The resulting semi-empirical model (described in Ref. 25) has been reported to provide accurate and efficient predictions of signal amplitude along arbitrary illuminated signal paths whenever Mode 1 remains the dominant mode of the signal propagating along the entire path.

*Sub-models are simple mathematical functions approximating the attenuating rate and excitation-factor amplitude of the Mode 1 signal; they are functions of the signal path's geophysical properties such as ground conductivity, geomagnetic latitude, angle between the geomagnetic field and the path, and day or night illumination condition.

B.1.2 Signal Phase Deviation Prediction Models

A numerical model approximating the IPP-predicted multimode signal phase behavior was used for predicting the modal interference-induced phase deviations in both 10.2 and 13.8 kHz signals along day and night paths. For mixed paths (where IPP is not applicable) different models were used for the two frequencies. The 10.2 kHz phase deviation information (developed five years before the 13.8 kHz phase information) was generated using a semi-empirical algorithm based on observed phase deviations in 10.2 kHz signals from the Liberia and Argentina stations. The observations indicated that severe "nighttime-type" modal interference effects did not generally occur on mixed paths which are more than two-thirds in daytime. Exceptions have been noted, of course, but these observations provided the basis for the semi-empirical algorithm used for estimating the phase deviations in the 10.2 kHz signals along mixed paths. The algorithm first notes whether the mixed-path coverage prediction point in question is subjected to excessive modal interference effects (i.e., $\Delta\phi > 20$ cec) when the entire prediction path is assumed to be under the nighttime illumination condition. If the nighttime interference effects are estimated to be excessive at the prediction point, and if the nighttime portion of the path is more than one-third of the total path length, the prediction point is assumed to have excessive mixed-path modal interference effects. Otherwise, the prediction point is considered to be non-modal (i.e., $\Delta\phi \leq 20$ cec).

In developing the 13.8 kHz phase deviation information, a more robust theoretical approach was used for predicting phase deviations along mixed paths. The approach combined the multimode signal amplitude and phase behaviors along the daytime and nighttime segments of a mixed path (determined using the theoretical IPP model, Ref. 20) with the mode conversion effects (obtained using the mode conversion routine of the theoretical FASTMC model, Ref. 24) to account for the exchange of mode energy at the path-terminator discontinuity along the mixed path (see Fig. A-5). A description of the resulting phase deviation prediction algorithm is given in Ref. 26.

B.2 ATMOSPHERIC NOISE AMPLITUDE PREDICTION MODEL

Electromagnetic atmospheric noise in the VLF band is generated mainly by lightning discharges associated with thunderstorms. The amplitude of this noise field at any particular receiver location on the earth's surface is a combination of electromagnetic fields radiated by thunderstorms worldwide. The best-known noise amplitude prediction model is the model described in CCIR Report 322 (Ref. 27), in which global noise measurement data are fit with smooth curves as functions of geographic latitude and longitude, time of day, season of the year, and frequency. Another approach to the problem of VLF noise field predictions has been carried out by Maxwell (Ref. 28), and modified by the U.S. Naval Research Laboratory. This modified model is believed to present an improvement over the CCIR model and was used to generate noise amplitudes. The computer program implementation of the modified model is given in Ref. 29.

REFERENCES

1. Doubt, R.J., "Omega Regional Validation Status Report," Proc. of the Seventh Annual Meeting, International Omega Association (Arlington, VA), October 1982).
2. May, W.K., "Omega: A System Operator's Perspective," Proc. of the Ninth Annual Meeting, International Omega Association (Seattle, Washington), August 1984.
3. Bortz, J.E., Sr., Gupta, R.R., Scull, D.C., and Morris, P.B., "Omega Signal Coverage Prediction," Navigation: Journal of the Institute of Navigation, Vol. 23, No. 1, Spring 1976.
4. Gupta, R.R., Donnelly, S.F., Morris, P.B., and Vence, R.L., Jr., "Omega Station 10.2 kHz Signal Coverage Prediction Diagrams," Navigation: Journal of the Institute of Navigation, Vol. 27, No. 2, Summer 1980.
5. Gupta, R.R., Donnelly, S.F., Morris, P.B., and Vence, R.L., Jr., "Omega System 10.2 kHz Signal Coverage Diagrams," Proc. of the Fifth Annual Meeting, International Omega Association (Bergen, Norway), August 1980.
6. Gupta, R.R., Donnelly, S.F., Creamer, P.M., and Sayer, S., "Omega Signal Coverage Prediction Diagrams for 10.2 kHz, Volume I: Technical Approach,"

- The Analytic Sciences Corporation, Technical Report TR-3077-1 (ADA 092741), October 1980.
7. Gupta, R.R., Donnelly, S.F., Creamer, P.M., and Sayer, S., "Omega Signal Coverage Prediction Diagrams for 10.2 kHz, Volume II: Individual Station Diagrams," The Analytic Sciences Corporation, Technical Report TR-3077-2 (ADA 092742), October 1980.
 8. Gupta, R.R., Donnelly, S.F., Creamer, P.M., and Sayer, S., "Omega Signal Coverage Prediction Diagrams for 10.2 kHz, Volume III: Composite Diagrams," The Analytic Sciences Corporation, Technical Report TR-3077-3 (ADA 092743), October 1980.
 9. "Omega Coverage Diagrams: 10.2 kHz," Defense Mapping Agency Hydrographic/Topographic Center, DMA Stock No. OMPUB224COVDIAG, 1983.
 10. Gupta, R.R., and Morris, P.B., "Omega Modal Interference Maps," Proc. of the Eight Annual Meeting, International Omega Association (Lisbon, Portugal), August 1983.
 11. Gupta, R.R., and Warren, R.S., "Omega Station 10.2 kHz Signal Selection Made Easy," Proc. of National Meeting of Aerospace Institute of Navigation (Trevose, PA), April 1981.
 12. Swanson, E.R., "A New Approach to Omega Coverage Diagrams," Proc. of the Eight Annual Meeting, International Omega Association (Lisbon, Portugal), July 1983.
 13. Swanson, E.R., "Omega," Proc. of IEEE, Vol 71, No. 10, October 1983.
 14. Swanson, E.R., "Omega Coverage Accuracy at Specified Times," Proc. of the Ninth Annual Meeting, International Omega Association (Seattle, Washington), August 1983.
 15. Swanson, E.R., Kugel, C.P., and Doubt, R.J., "Indian Ocean Validation," Proc. of the Tenth Annual Meeting, International Omega Association (Brighton, England), July 1985.
 16. Gupta, R.R., Morris, P.B., and Doubt, R.J., "Omega Signal Coverage Prediction Diagrams for 13.6 kHz," Proc. of the Tenth Annual Meeting, International Omega Association (Brighton, England), July 1985.
 17. Gupta, R.R., and Morris, P.B., "Assessment of Omega 13.6 kHz Signal Modal Interference," Proc. of the Ninth Annual Meeting, International Omega Association (Seattle, Washington), August 1984.
 18. Gupta, R.R., "Omega Signal Coverage Prediction Diagrams for 13.6 kHz," The Analytic Sciences Corporation, Technical Report TR-4418-6, August 1985.
 19. Warren, R.S., Tench, K.A., Gupta, R.R., and Morris, P.B., "Omega ACCESS: A Microcomputer Display of Omega Signal Coverage Diagrams," Proc. of the Forty-Second Annual Meeting, The Institute of Navigation (Seattle, Washington), June 1986. (Also Published in Proc. of the Eleventh Annual Meeting, International Omega Association (Quebec, Canada), August 1985.)
 20. Ferguson, J.A., "Report on the Integrated Prediction Program with Nuclear Environment (IPP-2)," NELC Technical note TN 1980, July 1971.
 21. Gupta, R.R., "Impact of Omega Station Outage on 10.2 kHz Signal Coverage," Proc. of the Seventh Annual Meeting, International Omega Association (Arlington, Virginia), October 1982.
 22. Rzonca, L., "Omega Transmitter Outages January to December 1979," Report No. FAA-RD-80-113/FAA-CT-80-196, FAA Technical Center, October 1980.
 23. Rzonca, L., "Worldwide Omega and Very Low Frequency (VLF) Transmitter Outages During January to December 1980," Report No. FAA-RD-81-28/FAA-CT-81-26, FAA Technical Center, May 1981.
 24. Ferguson, J.A., and Snyder, F.P., "Approximate VLF/LF Waveguide Mode Conversion Model, Computer Applications. FASIMC and BUMP," Naval Ocean Systems Center, NOSC Technical Document 400, November 1980.
 25. Reynolds, R.A., and Gupta, R.R., "Omega Signal Amplitude Model Computer Program Documentation," The Analytic Sciences Corporation, Technical Report TR-4418-5, August 1985.
 26. Gupta, R.R., "Omega Signal Phase-Deviation Algorithms Development," The Analytic Sciences Corporation, Technical Report TR-4418-2, March 1984.

OMEGA NAVIGATION SIGNAL CHARACTERISTICS

by

Peter B. Morris
 US Coast Guard,
 Omega Navigation System Center,
 7323 Telegraph Road
 Alexandria, Virginia
 United States

and

Radha R. Gupta
 The Analytic Sciences Corporation
 55 Walkers Brook Drive
 Reading, Massachusetts 01867
 United States

ABSTRACT

A description of Omega/VLF signal propagation is given. Particular emphasis is given to "non-standard" signal propagation scenarios including propagation over regions of low ground conductivity, signal spreading and converging, antipodal effects and long-path reception, modal interference (including fast terminator transit and off-path effects), and temporal anomalies (SIDs, PCAs, and magnetic storms). These elements of signal behavior are described qualitatively to aid in understanding the basis for signal selection algorithms employed in conventional Omega/VLF receiving systems. Equipped with this knowledge, the user may invoke manual deselection procedures when the receiver is suspected of processing an undesirable signal, i.e., one likely to produce significant navigational error. As further guidance, a table of recommended signal deselections is given for approximately 80 geographic locations around the earth. Signals are recommended for deselection on the basis of modal interference, long-path reception, and solar proton activity.

1. INTRODUCTION

The Omega Navigation System is a worldwide, all-weather, medium accuracy, very low frequency (VLF) radionavigation system transmitting formatted CW signals at 10.2, 11.05, 11.33 and 13.6 kHz from the eight transmitting stations located in seven nations. Besides navigation/positioning, the signals are used for timing and frequency control. A variety of Omega navigation receivers are available, ranging from simple units to very sophisticated systems integrated with other types of navigation equipment.

The Omega system user is usually only aware of the performance of the network of transmitting stations through the performance of his receiver. Thus, if the receiver malfunctions or suffers from poor design, it is frequently said that "Omega" has failed. Since user receiver interaction is important to the effective use of Omega, it is essential to acquaint the user with some basic characteristics of Omega signal propagation. Many of the peculiarities/irregularities of Omega signal behavior can be programmed into microprocessor-based receivers and thus be made transparent to the user. However, a surprising number of signal propagation characteristics are not or cannot be readily pre-programmed. In such cases, an informed user will be able to make the wisest decision regarding selection or deselection of signals.

The objective of this paper is to describe the basic characteristics of both normal and anomalous Omega signals as received in different parts of the world at various times of day. In Section 2, we briefly discuss the basic features of Omega signal propagation. The less common characteristics of Omega signal propagation are described in detail, including: propagation over low ground-conductivity regions, signal spreading and focusing, long-path reception, modal interference, and temporal anomalies.

Section 3 contains a table of expected Omega station signal deselections at about 80 geographical sites throughout the world. Rather than list "preferred" station signals or attempt to rank signal preference, we present the table as a deselection guide. The reason for this "negative" listing is that most currently-available receivers treat "normal" signals in the proper way - weighting the received signal phases by the respective signal-to-noise ratios and combining the signal phases from the stations having the best station-to-user geometry. However, abnormal signal behavior is not always detected or well-predicted by algorithms found in currently available receivers, and the presence of such undetected signals may have disastrous consequences for navigation/position-fix-

ing. Thus, a deselection guide is useful for reducing the incidence of calamitous errors rather than making marginal improvements to reasonable navigation accuracy.

2. SIGNAL PROPAGATION CHARACTERISTICS

2.1 Propagation of Vlf Waves Under Idealized Conditions

VLF electromagnetic waves are propagated to long distances (5000 to more than 20,000 kilometers) in the space between the earth and the so-called D- and E- regions of the ionosphere, called the "earth-ionosphere" waveguide. The long signal range is due to the generally low attenuation (loss) of the "guided" wave. The wave suffers little attenuation for three principal reasons:

- a) The lower boundary (i.e., the earth's surface) of the waveguide has generally high conductivity ($>10^{-3}$ mho/m) and, thus, waves do not easily penetrate the earth's surface.
- b) The earth's atmosphere (below the D-region) has extremely low conductivity and is effectively represented by a vacuum.
- c) The D- and E-regions of the ionosphere (80-150 km) have low average conductivity ($\sim 10^{-5}$ mho/m) but have a steep conductivity gradient between 70 and 100 km, which effectively reflect VLF waves.

The earth's magnetic field, acting upon ions in the D- and E-regions, makes the ionosphere a magnetized plasma. This introduces anisotropy, i.e., the properties of a wave interacting with the ionospheric plasma which differ markedly depending on the direction of propagation with respect to the geomagnetic field. Because of this anisotropy, a wave travelling from east to west will give up a large fraction of its energy to electrons and ions in the ionosphere (east-west effect); much less energy is lost to the ionospheric constituents in traveling from west to east.

The ionic structure of the ionosphere is quite sensitive to the net solar illumination incident upon it. During the day, solar photoionization maintains a small but stable ionized component between 70 and 80 km. At night, the absence of solar radiation and the high collision frequency between ions and neutral molecules cause the D-region to disappear almost entirely so that VLF waves are effectively reflected from the bottom of the E-layer (80-90 km).

This difference in ionospheric structure leads to differences in day and night propagation of waves between the earth and ionosphere boundaries. Since these boundaries are separated farther at night, a wave making oblique reflections from the earth and ionosphere will interact less with the night ionosphere than with the day ionosphere for the same lateral distance over the earth's surface. Thus, one would expect lower wave attenuation at night which is, in fact, observed.

Lower wave attenuation is, of course, a desirable feature, but a wider "waveguide" (the region between the earth and ionosphere) at night means that more than one waveguide-mode* may be present. In cases where these modes have comparable signal strength, the total-signal phase does not vary linearly with distance and radionavigation using these signals is not generally possible.

2.2 Low Ground-Conductivity Regions

As mentioned above, the earth's surface usually has a VLF ground conductivity greater than about 10^{-3} mho/m. However, in polar regions with large areas of fresh-water ice and tundra the conductivity is less than 10^{-3} mho/m, so that a significant fraction of the wave energy is absorbed. Greenland and Antarctica, in particular, are large land areas having VLF conductivities between 10^{-5} and

3×10^{-5} mho/m. Waveguide modes with electric field vectors nearly perpendicular to the earth's surface are most affected by these regions, over which the signal attenuation rate is as large as 30 dB/1000 km. This large attenuation rate at Omega frequencies is the source for the term "Greenland Shadow," especially in connection with the Norway Omega station.

*A discussion of the waveguide-modes is contained Section 2.4.

It should be noted that because of comparatively large skin depth (about 150 meters for a conductivity of 10^{-3} mho/m) of the wave relative to the earth's surface, ground conductivity at Omega frequencies is not sensitive to day, night, seasonal, or short-term weather conditions (such as snowfall). However, for a wave propagating between the ionosphere and this low conductivity region, the difference between day and night propagation is significant since the daytime waveguide is narrower and hence the wave interacts more frequently with the lossy earth's surface than at night.

2 3 Spherical Earth Effects

Radionavigation systems which operate at frequencies much higher than Omega frequencies are limited to ranges which permit approximation of the earth's surface as a flat plane. In these cases, the geometry of propagation (and also position-fixing algorithms) is relatively simple. However, because the ranges encountered in Omega signal propagation are on the order of the earth's radius, spherical-earth effects must be taken into account. Since a signal launched between the earth and ionosphere is constrained to propagate, for the most part, between these two boundaries and because the "width" of the waveguide (70-90 km) is much less than its "length" (thousands of kilometers), the signal may be viewed as a two-dimensional wave constrained to the surface of a sphere. In this view, a wave propagates outward from a VLF source with circular wave fronts (see Fig. 2.3-1). As they travel outward, these circles become larger and larger until they reach the "equivalent equator" (the equator corresponding to a north or south pole located at the source). At this point, the energy per unit length of the wavefront (due purely to geometry) is a minimum. Beyond the equivalent equator, the circular wavefronts diminish in size and that energy density grows until the wave converges at the antipode (pole opposite to that assumed for the source, or transmitter). This model can aid in understanding why stronger signals are sometimes received at more distant receiver locations. For example, Omega Japan station signals in Buenos Aires are observed to be much stronger than those received in Easter Island, Tahiti, or the West Coast of the United States. This is because the antipode of the Japan station is a short distance east of Buenos Aires in the South Atlantic Ocean and the signal is substantially "focused" there.

If the idealized case discussed above is modified to include the east-west effect, then we see that near the antipode signals arrive mainly from the west. Figure 2.3-2 shows the situation near the antipode where the only long-path signals are those propagating east through the antipode. The range attained by the long-path signal depends upon both the upper- and lower-boundary characteristics of the entire long-path waveguide. As discussed before, if the lower waveguide boundary has substantial portions with low conductivity, the long-path signal guided by this boundary is severely attenuated. As also mentioned previously, the signal attenuation due to the upper boundary is principally due to the geomagnetic field and the ionospheric illumination by the sun. We have already accounted for the geomagnetic field (see Fig. 2.3-2) so that we must discuss the effect of solar illumination. We saw before that the wider waveguide associated

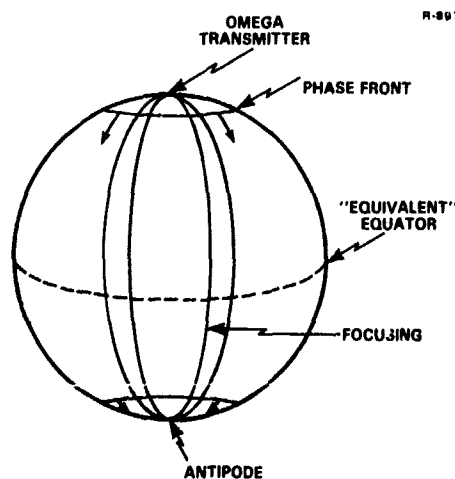


Figure 2.3-1

An Illustration Displaying Focusing of a Station's Transmitted Signals Beyond the Station's Equivalent Equator

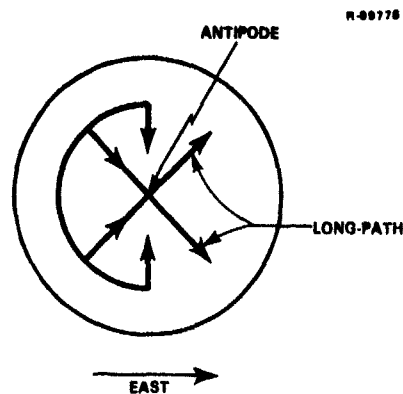


Figure 2.3-2 Semi-Circular Focusing Near a Station Antipode Due to the East-West Effect

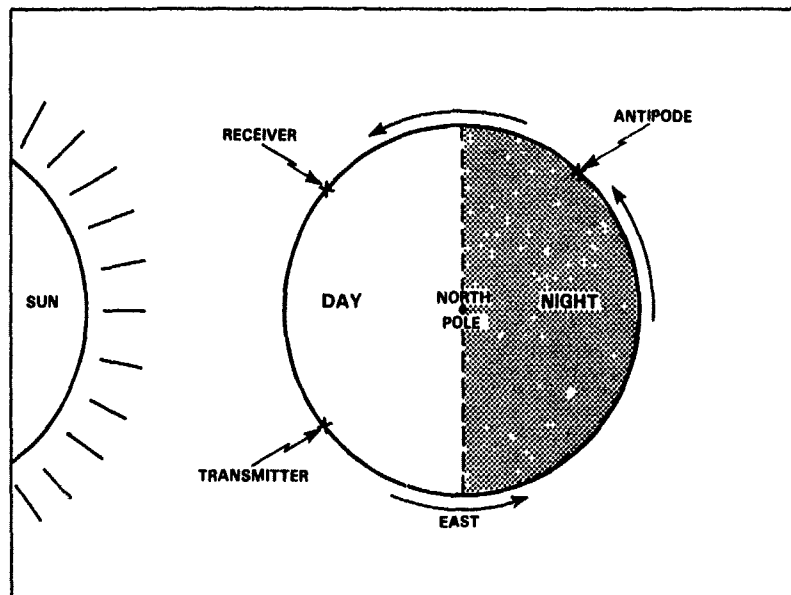


Figure 2.3-3 Ideal Long-Path Situation (Transmitter and Receiver are Assumed to Lie on the Equator, and Earth is at Equinox)

with the nighttime portion of propagation paths yields less signal loss than the narrower dayside portion of the propagation path. Thus, if the long-path (which, by definition, includes more than one hemisphere) subtends the entire nightside hemisphere, the signal attenuation is smallest and the long-path signal propagates farthest. On the other hand, if the long-path subtends the entire dayside hemisphere the attenuation of the long-path is maximum (see Fig. 2.3-3).

In summary, long-path signal reception from an Omega/VLF transmitting source is most likely to occur at a site when:

- The site is west of the source
- The long-path contains little or no low ground- conductivity portion(s)
- The long-path includes the entire night-time hemisphere, or equivalently, the short-path is all-day

2.4 Modal Characteristics

As discussed earlier, the earth and ionosphere act as waveguide boundaries for a VLF signal source located on the surface of the earth. As with other waveguide systems, one is interested in the modes (i.e., unique electromagnetic field patterns) of propagation permitted by the waveguide and the signal source. The literature associated with this problem is extensive (see, e.g., Refs. 1 through 12) so that this treatment only touches upon the main features.

Because the electromagnetic properties of the upper and lower boundaries of the waveguide along a path can change rather dramatically from point to point along the path, many researchers have treated the waveguide characterizing the path as being composed of many segments, each of which may be considered as a rectangular, two-dimensional waveguide with fixed upper and lower boundary properties. Earth curvature effects are included by mapping the index of refraction of the propagation medium as a function of geocentric radius (cylindrical geometry) onto the rectangular waveguide. Spherical-focusing effects are accounted for in an ad hoc manner by a multiplicative factor. Each component of the signal field is considered to be a sum of an infinite number of independent waveguide modes each of which is classed as either transverse-magnetic (TM) or transverse-electric (TE). A TM (or TE) mode is characterized by its magnetic field (or electric field) perpendicular to the direction of the propagation and to the local vertical. Excitation factors (complex quantities) are defined for each mode and describe the relative extent to which the transmitting source (or receiver) couples the mode energy into (or from) the waveguide.

Conventionally, the modes are numbered in order of increasing phase velocity, with TM modes odd-numbered and TE modes even-numbered. Thus, Mode 1 is the lowest phase-velocity TM mode, and Mode 2 is the lowest phase-velocity TE mode. At this point, we should note that Omega propagation correction algorithms* used in conventional receiver-processors are based on semi-empirical signal propagation models which inherently assume Mode 1 signal characteristics. Hence, Omega signals can be effectively used only if the total signal field is dominated by the Mode 1 component.

The signal fields in the vicinity of the VLF source are more complex (in terms of space and time variations) than the fields at significant distances (≥ 10 wavelengths) from the source. In the waveguide mode picture, this complexity is due to the presence of many modes near the source (see Fig. 2.4-1). The higher-order modes usually have a greater attenuation rate but also have a larger excitation factor (Ref. 1) so that, in general, these modes have substantial signal strength near the source but attenuate rapidly with distance, beyond 10-20 wavelengths from the source.

At night, the earth-ionosphere waveguide is wider, as noted earlier, and more modes are observed (i.e., those having comparable signal strength to each other and to Mode 1) than during the day. The higher-order modes are also less sensitive to the east-west effect than Mode 1. In some cases, these effects lead to the dominance of higher-order modes in westward propagation at night. The presence of higher-order modes with signal strength comparable to or exceeding that of Mode 1 is a condition known as modal interference.

As discussed previously, the influence of the east-west effect (as measured, for example, by the difference between east and west attenuation rates) depends directly on the size of the horizontal component of the geomagnetic field. Thus, at low geomagnetic latitudes (near the magnetic equator), east and west propagations are strongly differentiated. This further implies that Mode 1 signal amplitude is rapidly exceeded by higher-order modes at night as the reception point is moved westward from the transmitting source. If these so-called modal (i.e., mode interference) regions are "mapped-out", one finds that low-geomagnetic-latitude Omega transmitting stations are associated with much more extensive modal interference than are high-geomagnetic-latitude stations.

Because of the relatively long length of Omega/VLF propagation paths, it is quite probable that only a portion of the path is in daylight, leaving the other portion in the night hemisphere. In this case, the two-dimensional waveguide connecting the transmitter and receiver contains a physical discontinuity (as

*The propagation correction (PPC) is an estimate of the phase deviation from a nominal, or charted, phase value. The phase deviation depends upon signal path characteristics such as ground conductivity, magnetic dip angle, path bearing, solar zenith angle, etc.

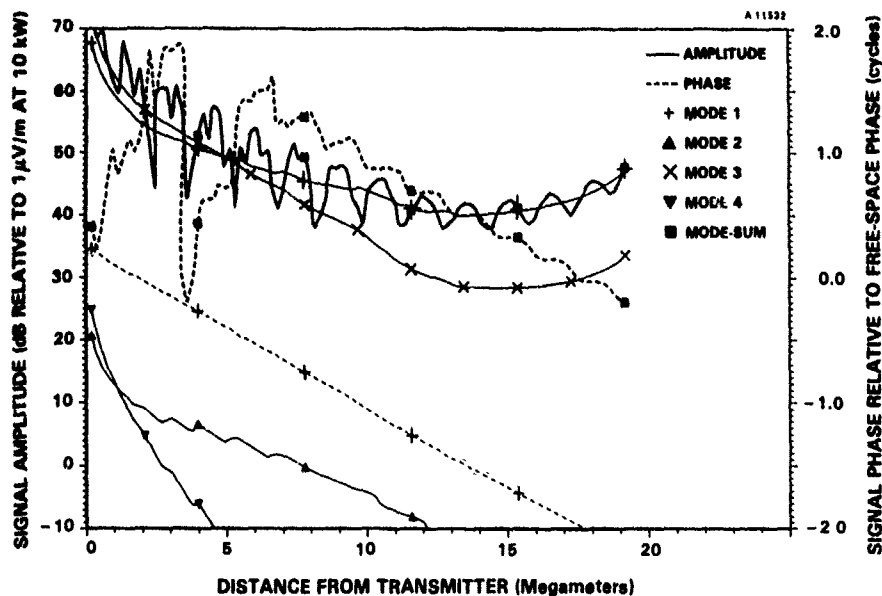


Figure 2.4-1 An Example of Multimode Omega Signal Along a Nighttime Path

contrasted to an electrical discontinuity at the boundary, e.g., Greenland or Antarctica). Since the waveguide-mode structure is different in the day and night portions, signal energy is said to be "converted" between allowable modes (see Fig. 2.4-2). Except in the vicinity of a transmitting source, a single mode is a good approximation to the multimode signal field during the day. At night, no more than three modes are usually needed to adequately characterize the multimode signal. Because the ionization dynamics differ markedly for sunrise and sunset in the D-region, the mode conversion differs for paths crossing sunrise and sunset terminators (day/night boundaries). The conversion also depends on whether the propagation is from day to night, or night to day.

Calculations of signal phase and amplitude using the waveguide-mode approach usually assume "all-day" or "all-night" path conditions (horizontally-homogeneous wave-guide). For paths crossing the terminator, mode conversion coefficients must be computed (normally a computationally burdensome task). Frequently, a user will seek guidance concerning the onset (disappearance) of "severe" modal

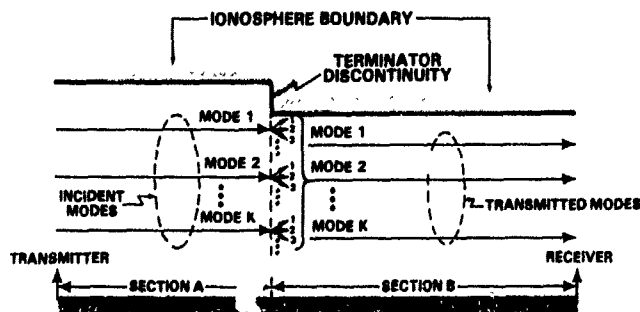


Figure 2.4-2 A Mixed-Path Waveguide Depicting Mode Conversion at the Path-Terminator Crossing

interference on a path transitioning from day to night (night to day). One very simplistic algorithm which has been used with apparent success is to first specify the position of the terminator with respect to a path which is known to be modal when fully dark. If the fractional portion of the path under the nighttime conditions exceeds some threshold (say, one-third) then the path is assumed to be corrupted by modal interference at that time.

One severe consequence of mode conversion for conventional Omega receivers is that under certain conditions modal dominance shifts (from Mode 1 to another mode) so quickly that the signal receiver (normally with a 3-5 minutes time-constant) may fail to track the corresponding sudden change in phase through the amplitude minimum and "slip" one or more signal phase cycles. If such a situation is suspected by a user, use of the signal should be discontinued until the path is no longer in modal dominance transition.

Mode conversion is said to be greatest (i.e., largest difference in mode structure across the path-terminator discontinuity, Ref. 13) when the path is normal to the terminator. As the path-terminator angle becomes smaller, the relative conversion of energy between modes across the (day-to-night) transition region decreases since the transition is now less abrupt. Although the mode conversion lessens, which is favorable to navigation, the speed at which the terminator crosses the path (or path crosses the terminator) increases. As the path-terminator angle approaches zero (path and terminator nearly parallel), the entire path may pass through transition during a time interval comparable to a receiver time constant. This effect may also cause temporary loss of signal track in the receiver system.

An additional effect associated with the small path-terminator angle configuration has been noted (Refs. 14 and 15). Observations indicate that signal phase changes occur even before the geometric terminator (defined by solar zenith angle) crosses a path or after it has completed crossing of the path. This has been explained by assuming "off-path" reflections occur from the approaching (or receding) terminator discontinuity. This effectively widens the transition zone* by approximately 10%.

2.5 Temporal Anomalies

In previous sections, we emphasized propagation-pertinent features of the earth-ionosphere waveguide which, though important, are relatively small when compared on a global scale. For example, the low ground-conductivity regions on the entire transition zone at any given time may occupy a comparatively small fraction of the earth's surface but affect Omega propagation paths to navigationally-important regions of the world. Analogous to these spatial anomalies are temporal anomalies which occur with durations very small compared to the average time between occurrences. However, when these anomalous events occur, they affect a substantial fraction of the entire globe and thus cannot be ignored. Here we will discuss three types of anomalous events of greatest significance to Omega/VLF signal propagation: Sudden Ionospheric Disturbance (SID), Polar Cap Absorption (PCA), and Geomagnetic Storm/Substorm.

Sudden Ionospheric Disturbance (SID) - A SID is caused by a solar flare occurring on the solar hemisphere facing the earth. The flares producing the most definitive SIDs are those with a substantial fraction of their energy in the X-ray spectrum. Flares are measured in terms of solar energy flux in the 1-8 Å band by earth-orbiting satellites. The two highest designated solar flare levels (and those which have the greatest impact on Omega/VLF navigation) are M-level (10^{-5} to 10^{-4} W/m²) and X-level (10^{-4} to 10^{-5} W/m²). The actual ionization produced by a flare in a region of the ionosphere associated with a point on the surface of the earth is indicated by the relative position of the sun; if the sun is directly overhead the effect is much greater than if the sun is low on the horizon. The strength of the effect is given by the cosine of the solar zenith angle (the angle of the sun with respect to the local zenith). Because of the long path-lengths associated with Omega/VLF propagation, a SID has an "integrated" effect on the signal at the point of reception; at each path segment, the phase perturbation is computed from the cosine of the local solar zenith angle and the total effect on the path is calculated as a sum of the perturbations at each path segment. This means that long paths passing through the sub-solar

*The day/night boundary is not a perfect step function. It is usually defined for solar zenith angles between 74° (84°) and 99° (99°) at 10.2 kHz (13.6 kHz), and hence describes a "transition zone" of finite width on the surface of the earth.

point exhibit maximum effect while short paths at local early morning (post-sunrise) and local late afternoon (pre-sunset) are affected the least. As an example, a path of average length (say, 7000 km) fully in the day hemisphere at low latitudes may easily show a 100 centicycles perturbation in the signal phase (at 10.2 kHz) during an X-level flare event.

Signal phase is normally advanced during a SID since the effective ionospheric reflection height is depressed due to increased ionization. The signal amplitude at the lower Omega frequencies is decreased during a SID, but above about 12.5 kHz the amplitude shows an increase. The effect of such an event on navigation is highly variable and quite sensitive to geographic location, time, and station signals received. For direct-ranging (i.e., rho-rho) navigation systems, the lines-of-position (LOPs) are effectively moved closer to the station as a result of this event. However, if signals are received from stations in opposite directions from the receiver and if some sort of averaging (e.g., least-squares, position-fixing algorithm) is employed, little position error may be caused by a SID. Hyperbolic navigation systems also may incur little position error if the station pairing (to form phase-difference LOPs) is done with well-correlated paths, i.e., paths whose integrated illumination conditions are similar.

SIDs normally exhibit (see Fig. 2.5-1) a rapid onset (5-10 minutes) to peak followed by a much slower (~45 minutes) recovery. Total durations vary widely - from less than 30 minutes to several hours. Because of the usually short duration, warning notices for these events are not published or broadcast in advance.

Polar Cap Absorption (PCA) - A PCA is said to occur when a solar flare event is accompanied by a large flux (above a certain threshold) of energetic protons at the earth. The net flux of these protons incident upon the earth depends significantly upon the earth-sun geometry. A PCA is most likely to occur when the proton-associated flare occurs on westerly longitudes of the solar disk facing the earth - including sources on the western limb and beyond. The proton flare giving rise to these events also usually (but not always) has a large X-ray component. Thus a SID frequently precedes a PCA. However, because protons are particles, they have arrival times varying from 15 minutes to some tens of hours depending on the source location on the sun and the topology of the interplanetary field (Ref. 16).

The protons emitted during a proton flare generally have a wide spectrum of energies: from a few thousands of electron volts (keV) to more than 50 million electron volts (MeV). The relative time duration for which these particles show

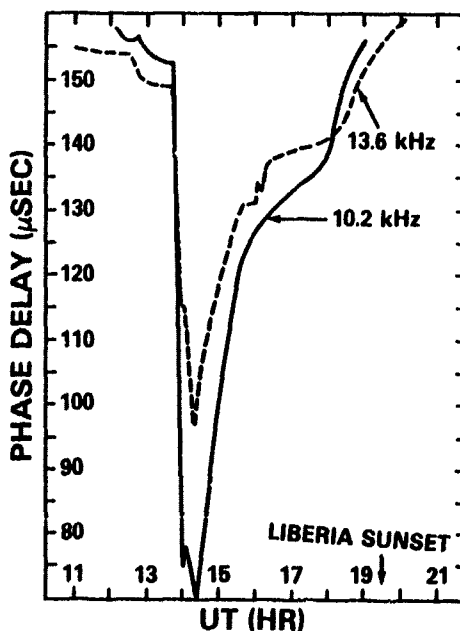


Figure 2.5-1 An Example of the SID Effect Observed on Omega Liberia Station Signals Received at Deal (NJ) (Ref. 20)

highly elevated flux levels is usually strongly dependent on energy. Typically, large fluxes of highly energetic protons (greater than 50 MeV) are observed for only a few hours whereas enhanced fluxes of the lower energy particles are known to persist for several days (Ref. 17). Upon reaching the earth, the high energy particles penetrate the ionosphere/atmosphere to low altitudes (~60 km and lower) but because of the high density of air molecules at these altitudes, the ion-pairs created quickly recombined, thus having little effect on VLF propagation. The particles from the very low energy part of the proton flare spectrum lose energy (due to collisions and scattering) at relatively high altitudes (ionospheric E- and F-regions) and do not contribute to the ionization of the D-region). Thus, the particles from the lower-middle range of the proton energy spectrum are the source of most of the excess ionization found in the D-layer. A recent study (Ref. 17) discovered that Omega signal phase advances on transpolar paths are highly time-correlated with the 6 MeV differential proton flux channel on the GEOS satellites.

These large differences in the persistence of elevated fluxes in the various components of the proton flare spectrum lead to some confusion in event warning. The definition of PCA used by the NOAA-USAF* Space Environment Services Center is that the cosmic noise absorption (at 30 MHz) as measured by a riometer inside the polar cap region (usually in Greenland) be greater than 2 dB in day time and 0.5 dB at night. Currently, however, the absorption is inferred from satellite measurements of proton flux at energies greater than 10 MeV. With this definition, PCAs generally last only a few hours while phase disturbances at VLF last for several days, as discussed above (see Fig. 2.5-2). For this reason the Omega Navigation System Center maintains an alerting system through continuous monitoring of transpolar paths between Omega transmitting stations. The Omega Hawaii to Omega Norway baseline path is especially well-calibrated and serves as a primary indicator of PCA conditions. Other paths which are monitored for PCA events include the Omega Japan to Omega North Dakota baseline and the Omega Argentina to Omega Australia baseline. A PCA is declared in effect if the Hawaii-to-Norway path phase is advanced by at least 20 centicycles for a sufficiently long period (roughly 2 hours) and if confirmed by other path observations and satellite proton flux data. Thus, it is recommended that Omega/VLF users heed the warnings broadcast by radio stations WWV/WWVH and transmitted on various alert circuits throughout the world.

Although the sources of ionization are different for SIDs and PCAs, both result in Omega/VLF phase advances for the regions affected. A typical PCA produces phase advances of approximately 40 centicycles over the Hawaii-Norway path (transversing roughly 8700 km of the geomagnetic polar region) although phase advances of 100 centicycles or more have been observed. In the direct-ranging navigation mode, PCAs often yield larger position errors than SIDs, since only one or two paths are usually affected and so averages are not taken over compensating error conditions. For receiver locations inside the polar-cap re-

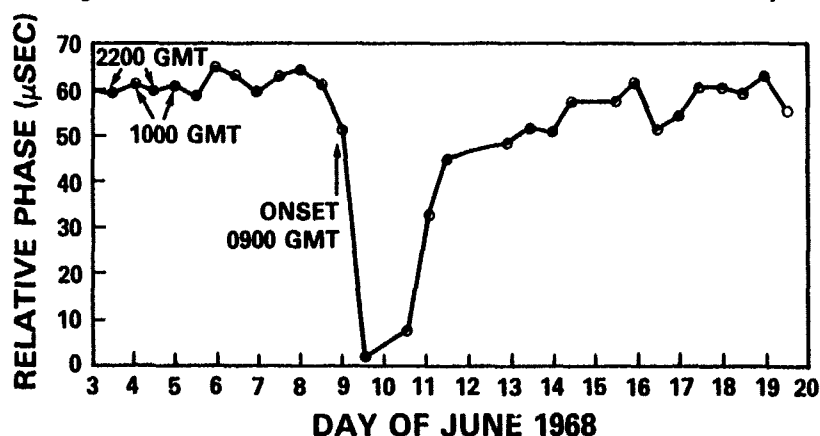


Figure 2.5-2 An Example of PCA Effect Observed on Omega Norway Station Signal Received at Hawaii (Ref. 20)

*NOAA-USAF: National Oceanic and Atmospheric Administration, U.S. Air Force.

† WWV (Ft. Collins, Colorado) and WWVH (Kauai, Hawaii) transmit time information, geophysical alerts, weather advisories, and Omega status.

gion, the error situation is similar to that for a SID. In the hyperbolic navigation mode, path correlation cannot be advantageously used, except in the case where two paths pass through roughly equal sections of the polar region. In this as well as for a SID events, of course, frequency-difference schemes may be effectively used to significantly reduce the Omega position-fix errors.

Geomagnetic Storms - Geomagnetic Storms (or substorms) occur in response to solar dynamical processes which rapidly modify the magnetic field structure near the surface of the sun. This field disturbance is propagated outward from the sun along interplanetary magnetic field lines. When the disturbance reaches the near space environment of the earth ("geospace"), the geomagnetic field is similarly affected due to its connective relationship to the interplanetary field. The importance of the disturbance (normally a compression of the field) is not so much due to the changed geomagnetic field at a given location (rarely more than 1% of the total geomagnetic field) but rather to the precipitation of electrons from the magnetosphere caused by the sudden change in field structure. Normally, a sizeable population of electrons and ions is "trapped" in the middle magnetosphere. These particles bounce back and forth along geomagnetic field lines between "mirror points" where the increased magnetic field strength (due to converging field lines) prevents further penetration. When the field line structure is suddenly changed, electrons are "dumped" into the ionosphere at the base of the magnetic field lines. This electron precipitation occurs mainly in the auroral zone which lies roughly between geomagnetic latitudes of 60° and 65°. However, as mentioned above, the geomagnetic field is generally compressed during magnetic storms and in the extreme cases the auroral zone, which is tied to the geomagnetic field structure, may move equator-ward by as much as 20° in geomagnetic latitude.

As in the case of SIDs and PCAs, the phase of Omega/VLF signals is advanced as a result of the increased ionization caused by the precipitation of electrons. However, because the auroral zone is rather limited in extent (~500 km), paths which are roughly perpendicular to the auroral "band" show a relatively small effect from a storm/precipitation event. Paths which are "parallel" to and lie within the auroral zone for a substantial fraction of the path length do exhibit a sizable effect, however. A major magnetic storm can have a secondary effect as, for example, in the case of a PCA (which are frequently accompanied by magnetic storms) during which the effective polar cap region is expanded due to the equator-ward movement of the auroral zone, resulting in a greater fraction of paths affected by the PCA.

3. SIGNAL CHARACTERISTICS: APPLICATIONS

We now turn from a discussion of general Omega/VLF wave propagation characteristics to specific applications involving these characteristics. As mentioned in Section 1, guidance for the appropriate signal mix for use in Omega navigation/position-fix is presented from a "negative" viewpoint, i.e., signal deselection as opposed to signal selection. Two reasons motivate this presentation. First, most Omega receiver-processors employ efficient algorithms for selecting which signals to use (or better, the weights associated with each signal) in the position/navigation tracking filter. Thus, little additional guidance is needed there. However, signal deselection algorithms coded into conventional receiver-processors may be based on limited external data and thus not generally applicable. Second, it is difficult to "rank" most good signals a priori since signal-to-noise ratio predictions do not include local, artificial sources of noise and because signal combinations are often more important to navigation than individual signals.

For the above reasons, we present in Table 3-1 an Omega station signal deselection guide for 79 geographic locations on the globe. For ease of reference, the site locations are ordered by latitude interval: arctic, northern mid-latitude, equatorial, southern mid-latitude, and antarctic. Within each latitude category, geographic location/site names (usually cities) are given alphabetically. Coordinates of the site locations are given to the nearest degree to emphasize the validity of these deselections over a reasonably-sized area.

Signals indicated as "polar" represent paths whose midpath portion* intersects the geomagnetic polar region (absolute value of geomagnetic latitude greater than 60°). During a PCA, these paths should be deselected. For a user inside the polar regions, the discussion of Section 2 applies, i.e., less navigational error should be incurred (given the proper mix of signals) than for users outside the region. It has been suggested that receiver-processor algorithms be developed to "de-weight" rather than deselect transpolar paths during PCAs. Such receivers would run less risk of signal "starvation" and, in fact, should track well upon passage into or out of disturbed polar regions.

The second deselection column in Table 3-1 lists signals recommended for deselection due to nighttime (where the entire path is assumed dark) modal interference. This includes all types of modal interference (discussed in Section 2) producing phase deviation (absolute difference between Mode 1 phase and observed phase) greater than about $1/5$ of a cycle. The nighttime modal interference guidance raises the question of what signals are recommended for deselection (due to modal interference) if the paths are not fully dark? Without specific signal coverage guidance (currently, coverage diagrams are provided for only two global times and four different months, Refs. 18 and 19), a deselection doctrine should be selected which best serves particular needs. For example, if signals are plentiful or one prefers a conservative approach, then the nighttime modal interference deselections can be made whenever any fraction of the path is dark. On the other hand, if signals are scarce or a riskier approach is acceptable, one may choose to deselect these signals only when the entire path is dark. Alternatively, a particular fraction of path darkness (e.g., one-third) may be used as a threshold, as discussed earlier.

The third column of recommended deselections in Table 3-1 covers those situations in which the short-path signals can never be used due to modal interference during path darkness, and the presence of long-path signals when the short-path is in daylight. For those times when the path is in transition, the indicated signal is likely to be dominated by a higher-order mode, comprised of several modes of comparable strength, arriving via the long-path, or a combination of long- and short-path signals.

It should be recognized that deselections are not differentiated by frequency (e.g., 10.2, 13.8 kHz), i.e., if either frequency of a station signal is problematic, the station signal is recommended for deselection. Also, the deselections shown are based partly on theoretical calculations and partly on observations. Some path deselections are inferred from observed/computed paths having similar characteristics, e.g., path length, geomagnetic field orientation, ground conductivity, and illumination condition.

Figure 3-1 gives a map of the geographic locations listed in Table 3-1 keyed by entry number. Omega stations are indicated by open triangles and keyed by station letter designations.

4. CONCLUSIONS

We have attempted to provide guidance to the users of Omega/VLF signals by describing, in general terms, the characteristics and peculiarities of these signals. This background should provide a better understanding of the basis for the algorithms used in conventional receiver-processors. The discussion of signal behavior and the tabulated deselections furnish an adequate basis for making intelligent manual station deselection when necessary.

*That part of the path interacting with the ionosphere, i.e., excluding ~ 850 km portion surrounding the transmitting source and receiver.

†Most Omega receiver-processor algorithms assume that the received multimode Omega signal is adequately approximated by the signal's Mode 1 component.

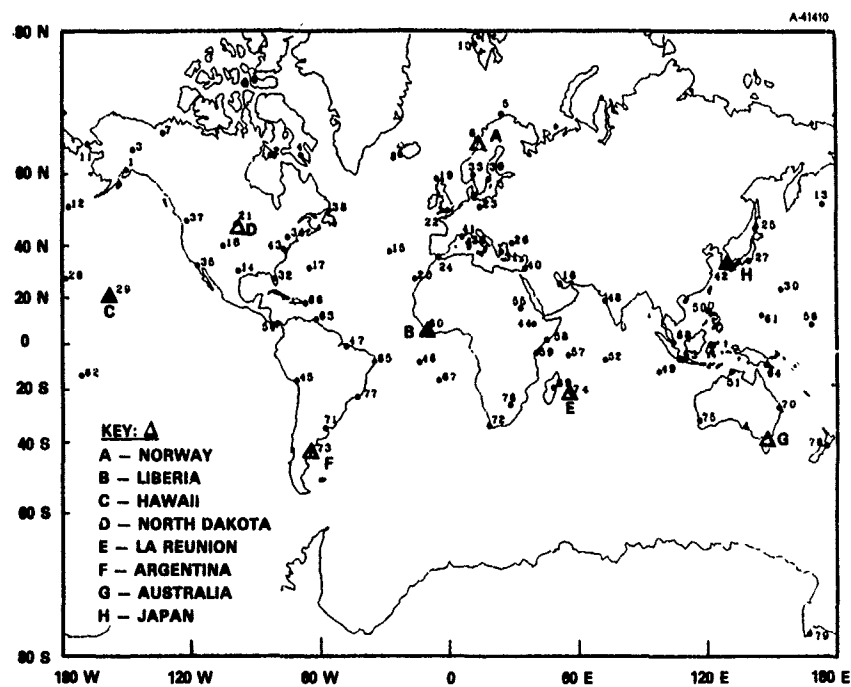


Figure 3-1 Geographic Locations of the Sites
Contained in Table 3-1

TABLE 3-1
OMEGA STATION DESELECTION CHART

REGION ¹	LOCATION	GEOGRAPHIC		RECOMMENDED DESELECTIONS		
		LAT (deg)	LONG (deg)	POLAR ²	SHORTLINE MODAL	SHORTLINE MODAL/ DAYTIME LONGPATE
Arctic (LAT > 60° N)	1 Anchorage, AK, USA	61	-150	A,B,D,E		F
	2 Coral Harbor, Canada	64	-83	A,B,C,D,E,F,G,H	F	
	3 Fairbanks, AK, USA	65	-147	A,D,E,H		E,F
	4 Prosser Bay, NWT Canada	64	-69	A,C,D,E,G,H	F	
	5 Hammerfest, Norway	71	24	C,D,G,H		A
	6 Neustad, Norway	67	13	C,D,G,H		E
	7 Inuvik, Canada	68	-133	A,B,D,E,F,H	B	
	8 Keflavik, Iceland	64	-23	C,D,H		A,E
	9 Resolute Bay, NWT Canada	75	-95	A,B,C,D,E,F,G,H	F	
	10 Spitzbergen, Norway	79	12	C,D,E,F,G,H	A	
	11 Vello, AK, USA	66	-168	A,B,D,E		F
Northern Mid-Latitude (20° S-LAT < 60° N)	12 Adak, AK, USA	52	-177	A,B		F
	13 Attu, AK, USA	53	173	A,B		F
	14 Austin, TX, USA	31	-98		B,F	A,E
	15 Azores, Portugal	38	-28	C		E,G,H
	16 Abu Basa, Bahrain	26	51	D	H	
	17 Bermuda Is., UK	32	-65	E	B	E
	18 Boulder, CO, USA	40	-105	A,E	D	
	19 Butt of Lewis, UK	59	4	C,D,H	A	
	20 Canary Is.	28	-16	C,G,H		E
	21 Dickey, MD, USA	47	-69	A,H	B	E,F
	22 Farborough, UK	51	-1	C,D	A	E,F,H
	23 Frankfurt, FRG	52	14	A,C,D	E,H	
	24 Gibraltar	36	-5	C		E,G,H
	25 Hakodate, Japan	46	142	A,B,D	H	
	26 Istanbul, Turkey	41	29	C,D	C,H	O
	27 Izu-Oshima, Japan	35	139	A,D	C,H	
	28 Kure Is., USA	28	-178	A	C	B,D,F
	29 Nagasaki Pt., Oahu, HI, USA	21	-158	A		F
	30 Narita Is., Japan	36	140	A,B	C	G
	31 Nos Maki, Greece	38	24	C,D	E,H	
	32 Orlando, FL, USA	28	-81	A,H	B,F	E
	33 Oslo, Norway	60	11	C,D,H	A	C
	34 Rome, IT, USA	43	-15	A,H	D	
	35 San Diego, CA, USA	33	-117	A,E		B,E,F
	36 Sardinia, Is., Italy	41	9	C,D	E	C,H
	37 Seattle, WA, USA	48	122		D	A,B,E,F
	38 St. Anthony, Newfoundland	51	-54	C,D,G,H	B,E	
	39 Stockholm, Sweden	59	18	C,D		A,G
	40 Tel Aviv, Israel	32	35	C,D	B,C,H	O
	41 Toulon, France	43	6	C,D	E	C,H
	42 Tsushima, Japan	34	129	A,D	C	F
	43 Washington, DC, USA	39	-77	A,H		B,D,F
Equatorial (20° S-LAT < 20° N)	44 Addis Ababa, Ethiopia	9	39	D		B,C,G,H
	45 Acre, Peru	-10	-71	H	B,F	E
	46 Ascension, Is., UK	8	-14	C	A,B,E	
	47 Belm, Brazil	-1	-48	G,H	B	
	48 Bombay, India	19	72	D	C,H	C
	49 Cocos Is., Australia	-12	97	D	C,H	
	50 Cubi Pt., Philippines	15	120	A,D,F	C,G,H	
	51 Darwin, Australia	-12	131	A,F	C,H	
	52 Diego Garcia Is., UK	-7	72	D	E,H	
	53 Djakarta, Indonesia	-6	107		C,E,G,H	D,F
	54 Galea Is., Panama	9	-80	A	F	B,E
	55 Khartoum, Sudan	16	33	C,D	E	
	56 Kwajalein, Micronesia	9	168	A,B	C	
	57 Kato Is., Seychelles	-5	55		E,H	C,D
	58 Mogadishu, Somalia	2	45	D	E,G	
	59 Mombasa, Kenya	-4	40	D	E,G	
	60 Monrovia, Liberia	6	-11		E	C,H
	61 Orkney, Guam	13	145	A	C	
	62 Pago Pago, American Samoa	-14	171	A	C	F
	63 Piarco, Trinidad	11	62	H	B,F	E
	64 Port Moresby, Papua New Guinea	-9	147	A	C	
Southern Mid-Latitude (20° S-LAT < 60° S)	65 Recife, Brazil	-8	-35	C,H	A,B	E
	66 Sabana Seca, Puerto Rico	18	-67	H	B,F	E
	67 Saint Helena Is., UK	16	-5	C	A,B,E	
	68 Singapore	1	104	D	C,G,H	
	69 Tananarive, Madagascar	-19	48	D		C,E
	70 Brisbane, Australia	-27	153	A,B	C,G	F
	71 Buenos Aires, Argentina	-35	-58	G	B,E,F	A
	72 Cape Town, South Africa	-34	18			C,H
Antarctica (LAT < 60° S)	73 Golfo Nuevo, Argentina	-43	-65	G	B	
	74 La Reunion Is., France	-21	54	D		H
	75 Perth, Australia	-32	116	F	C,E	
	76 Pretoria, South Africa	-26	28	G	E	O,H
	77 Rio de Janeiro, Brazil	-23	-43	C,H	A,B,E	
	78 Wellington, New Zealand	-41	175	A,E	C	B,D
	79 McMurdo, USA	-78	167	A,B,C,D,E,F,G,H	E	

¹ Deselect during PCA

² Note that "LAT" denotes the geomagnetic latitude

REFERENCES

1. Galejs, J., Terrestrial Propagation of Long Electromagnetic Waves, Pergamon Press, Oxford, 1972.
2. Wait, J.R., Electromagnetic Waves in Stratified Media, Pergamon Press, Oxford, Second Edition, 1970.
3. Watt, A.D., VLF Radio Engineering, Pergamon Press, Oxford, 1967.
4. Bickel, J.E., Ferguson, J.A., and Stanley, C.V., "Experimental Observation of Magnetic Field Effects on VLF Propagation at Night," Radio Science, Vol. 5, No. 1, January 1970.
5. Swanson, E.R., "Omega Coverage in India: A Case Study," Naval Electronics Laboratory Center, NELC Technical Report 1974, 15 January 1976.
6. Reder, F., Propagation Effects on Omega Signals and Methods of Reducing Them," Proceedings of the Fourth Annual Meeting of the International Omega Association (San Diego, CA), September 1979.
7. Gupta, R.R., and Morris, P.B., "Assessment of OMEGA 13.6 kHz Signal Modal Interference," Proceedings of the Ninth Annual Meeting of the International Omega Association (Seattle, WA), August 1984.
8. Burgess, B., and Walker, D., "Effects on Omega from Propagation Variations," Navigation: Journal of Institute of Navigation, Vol. 23, No. 1, 1970.
9. Crombie, D.D., "Periodic Fading of VLF Signals Received Over Long Paths During Sunrise and Sunset," Radio Science Journal of Research NBS/USNC-URSI, Vol. 68D, No. 1,
10. Walker, D., "Phase Steps and Amplitude Fading of VLF Signals at Dawn and Dusk," Radio Science Journal of Research NBS/USNC-URSI, Vol. 69D, No. 11, November 1965.
11. Gupta, R.R., and Morris, P.B., "Overview of OMEGA Signal Coverage," Navigation: Journal of the Institute of Navigation, Vol. 33, No. 3, Fall 1986.
12. Pappert, R.A., and Snyder, F.P., "Some Results of a Mode-Conversion Program for VLF," Radio Science, Vol. 7, No. 10, October 1972.
13. Lynn, K.J.W., "VLF Mode Conversion Observed at Middle Latitudes," Journal of Atmospheric and Terrestrial Physics, Vol. 35, 1973.
14. Mannheimer, D., "Lateral Bendix Effects at the Ionospheric Height Transition," Ionospheric Effects Symposium (NRL), 1981.
15. Mannheimer, D., "Twilight Bending Effects on Omega Navigation," Proceedings of the Eighth Annual Meeting of the International Omega Association (Lisbon, Portugal), July 1983.
16. Sauer, H., Private Communication, 1986.
17. Sauer, H., Spjeldvik, W., and Steele, K., "Omega Long-term Phase Advances," NOAA Technical Memorandum, ERL SEL-73, 1973.
18. Gupta, R.R., and Doubt, R.J., "OMEGA Signal Coverage Prediction Diagrams for 13.6 kHz," Proceedings of the Ninth Annual Meeting of the International Omega Association (Seattle, WA), August 1984.
19. Gupta, R.R., Donnelly, S.F., and Vence, R.L., "OMEGA Station 13.6 kHz Signal Coverage Prediction Diagrams," Proceedings of the Fifth Annual Meeting of the International Omega Association (Berger, Norway), August 1980.
20. Reder, F., "Propagation Effects on Omega Signals and Methods of Reducing Them," Proceedings of the Fourth Annual Meeting of the International Omega Association (San Diego, CA), September 1979.

POINTING CONTROL SYSTEM FOR THE TEAL RUBY EXPERIMENT

by

Ron Rogers and Monte Schlessinger
The Aerospace Corporation
P.O. Box 92957
Los Angeles, California 90009
United States

SUMMARY

Teal Ruby is an advanced earth-orbiting sensor that was developed to demonstrate the ability to detect airborne vehicles from space by use of infrared mosaic technology. The pointing system, which is able to direct the sensor at any designated target within the pointing envelope constraints and hold it steady with very low drift and jitter, represents a substantial increase in capability over earlier spaceborne systems. This pointing system, and the approach used to derive its performance requirements from the mission objectives, are the subject of this paper.

INTRODUCTION

The Teal Ruby sensor and spacecraft were developed under contract to DARPA and USAF Space Division as part of an ongoing infrared (IR) technology development program to provide a data base on scene backgrounds and airborne vehicles as seen from low earth orbit. The sensor comprises a cryogenically cooled infrared mosaic array mounted at the focal plane of a compound Cassegrainian telescope. The sensor enclosure and much of the associated electronics are gimballed such that they can move relative to the AFP-888 spacecraft, which maintains a locally level attitude. Angular freedom is such that the accessible field of view extends forward above the earth horizon, and aft approximately to the horizon. Lateral motion is sufficient to acquire and track objects up to 120 km either side of the satellite ground track. Figure 1 shows the system as it would appear in flight with the sensor acquiring data.

Many of the planned missions will involve cooperative ground targets. This requires extensive advance planning and demands a high degree of autonomy of the pointing system, since the missions must be uploaded and stored in advance, and then executed by the sensor pointing system at exactly the right time. Further, acquisition of high quality scene data from the staring array imposes stringent demands on pointing accuracy, drift rate, and line-of-sight jitter. It should be noted that sensor pointing is completely open loop; that is, pointing relies solely on navigation and the accuracy with which the various system components are aligned. This approach was necessitated by the diversity of targets and background scenes from which data must be acquired.

PROGRAM BACKGROUND AND OBJECTIVES

Prior to the initiation of the Teal Ruby space experiment activity, the DARPA Space Technology Office program had successfully demonstrated IR and visible mosaic detector arrays in a laboratory environment. There was considerable uncertainty associated with the practical use of this technology from space because target signatures are orders of magnitude more faint and spectrally, spatially, and temporally more complex than any measurement previously attempted. DARPA-sponsored system studies demonstrated that this new technology would provide the basis for considerable performance growth in space-based surveillance capabilities. In parallel with the technology development, an extensive field measurements program demonstrated that strategic aircraft, operating under cruise conditions, would produce contrast signatures above detection thresholds of this new technology when viewed from space.

In selecting a useful space-based experiment as a logical step to follow laboratory and ground-based experiments, DARPA chose aircraft detection as an illustrative stressing case to provide a meaningful demonstration of this technology. Thus, a practical quantitative "handle" on the potential utility of mosaic technologies would be gained for space-based surveillance applications for a wide range of potential operational missions. By performing a representative surveillance experiment, mosaic staring sensor technologies could demonstrate effectiveness for military space operations.

The overall objectives of the Teal Ruby Experiment (TRE) are to:

- o Provide a proof-of-concept demonstration of step-stare techniques using a mosaic focal plane to perform detection from space of subsonic strategic aircraft.
- o Collect multispectral target and background data from a space platform to provide a basis for the design of future operational IR surveillance systems.

SENSOR TRACKING REQUIREMENTS

In this section we discuss the origin of the performance specifications for the Teal Ruby pointing system and gimbal servos as derived from overall sensor performance requirements.

Gimbal Angle and Rate Requirements

For a satellite altitude of 380 nautical miles, the earth horizon angle is 62 degrees from the nadir. To view the earth's limb and above the horizon, the extent of the intrack gimbal is taken as +80 to -60 degrees. The remaining field of view from -60 to -90 degrees is blocked by the sensor stow cover.

The crosstrack angular freedom of +10 degrees is required for accessibility of points at least 120 km off the ground track of the satellite. It also permits crosstrack aircraft encounter experiments that reduce the overall position error ellipsoid. Larger crosstrack angles are impractical due to geometrically induced smear rates near the edge of the field of view.

The gimbal servos must be capable of driving the gimbals as required to point the sensor at any earth-fixed point in this field of view. This results in maximum velocity of 0.9 deg/sec for the intrack axis, with maximum acceleration of 0.0035 deg/sec² and crosstrack rate and acceleration maxima of 0.05 deg/sec and 0.0009 deg/sec² for all possible stare conditions. To provide additional capability for other than aircraft targets, the crosstrack rate and acceleration requirements were made equal to the intrack.

Line of Sight Accuracy Requirements

Allowable sensor pointing position error is fixed by target encounter considerations. The results of a single aircraft encounter with the aircraft flying crosstrack are summarized in Figure 2. The analysis, based on reasonable assumptions for aircraft navigation and spacecraft ephemeris accuracy, was for nadir passage and represents the most stressing condition.

The error ellipses shown represent the RSS of all error sources including sensor pointing accuracy, spacecraft ephemeris errors and target navigation errors. In the figure the target is shown across a column of three chips. The line of sight (LOS) position requirement is set to yield a single encounter probability on the order of 0.97. This translates into a pointing error requirement of 100 arc seconds one sigma. This resulted in the system pointing error requirement for each axis of 200 arc seconds with 0.95 probability.

Line of Sight Drift and Jitter Requirement

The design performance of the Teal Ruby Experiment sensor was established in terms of a parameter called the system equivalent target (SET). The SET is that theoretical target intensity wherein the signal-to-noise ratio is unity. Let us define the noise sources and their constituents.

SET (System Equivalent Target) - the total noise produced by the sensor (detectors, electronics, readout, etc.), the background and the clutter noise induced by the sensor drift and jitter.

CET (Clutter Equivalent Target) - the noise induced by the sensor line of sight drifting over a spatially structured background.

JET (Jitter Equivalent Target) - the noise induced by the sensor line of sight jittering over a spatially structured background.

NET (Noise Equivalent Target) - the noise produced by the sensor and shot noise from the background.

Assuming these components of the noise to be uncorrelated, we can write the following expression for the SET

$$SET = (NET^2 + CET^2 + JET^2)^{1/2}$$

The sensor NET sets the lower bounds on the system noise. The NET is a function of the optics area and efficiency, the sensor-to-target range, and the detector responsivity and integration time. Since we are discussing the control system requirements, only those noise sources (CET and JET) that are affected by the control system performance will be treated.

The CET and the JET are strong functions of the line of sight motion, the detector footprint, the power spectral density of the background and the signal processing technique. The calculation of the CET and JET are involved processes best suited to computer solution (Ref. 1). For the case of the clutter noise induced by drift (CET), the integral relationship that must be determined can be written as:

$$CET^2 = 2(L^2 \Delta \lambda) \int_0^\infty S(W) \cdot H_0^2(W) \cdot H_d^2(W) \cdot H_1^2(W) \cdot H_N^2(W) dW$$

where

$s(W)$ = power spectral density of the background

$H_g^2(W)$ = magnitude squared of the optics transfer function

$H_d^2(W)$ = magnitude squared of the detector transfer function

$H_i^2(W)$ = magnitude squared of the integration transfer function

$H_N^2(W)$ = magnitude squared of the differencing transfer function (assumes an N^{th} order background rejection filter)

W = temporal frequency

L = detector footprint size (projected in background)

$\Delta\lambda$ = spectral bandwidth

The calculation of the JET is somewhat more complicated because it not only involves the power spectral density of the background, but also the power spectral density of the jitter spectrum. For the TRE, an equivalent jitter amplitude was specified that accounts for both continuous and discrete jitter source due to such components as the spacecraft solar array motion, the earth shield vibrations, the TRE bearings and gyros. The expression developed by J. Rapier of Rockwell International (Ref. 2) to account for these effects is given as:

$$\sigma_j^2 = \left\{ \frac{2N}{\pi} \int_0^\infty \sum_j H_i^2(W) \left[P_j(W) + \pi \sum_i A_{ji}(W_{ji}) \delta(W - W_{ji}) \right] \frac{\sin^{2N+2} \left(\frac{WT}{2} \right)}{\left(\frac{WT}{2} \right)^2} dW \right\}$$

$H_j(W)$ = transfer function of j^{th} source to LOS

$P_j(W)$ = continuous power spectral density of j^{th} source

$A_{ji}(W_{ji})$ = amplitude of i^{th} discrete source

τ = frame time (look time)

σ_j = effective rms jitter amplitude

When formulated in this manner, the expression for the SET can be used to determine allowable pointing system drift rate and jitter. Assuming that the NET is fixed, various jitter levels are assumed, and the SET is calculated as a function of drift rate. When this is plotted along with expected target signal strength, it becomes apparent by inspection which combinations of drift rate and jitter will give acceptable signal-to-noise ratio (SNR).

Such a plot for four assumed jitter levels and two representative targets is shown in Figure 3. Note that there are two families of curves for the SET: one for a first difference signal processing algorithm, and one for second differencing. The target signal levels have been divided by four, since that was the desired SNR in the example. For small drift rates, the combined sensor NET and JET dominate, so the SET curves are horizontal. As drift rate increases, however, a point is reached where it becomes significant and the curves begin to slope. For high drift rates, the CET dominates and all curves approach an asymptote.

As long as the SET curve for a given jitter remains below a target line, that target is detectable. Thus, in the figure for a jitter level of 3.2 arc sec, neither target will be detectable using either first or second order differencing. For second order differencing and a jitter level of 0.1 arc sec, both targets are observable for drift rates up to approximately 50 M/s, etc.

For the Teal Ruby system, it was determined that a jitter force of 0.32 arc sec rms was attainable. A stringent requirement was imposed in the control system drift (rate error) in order to permit the best sensitivity in the vicinity of nadir to make possible optimum background temporal measurements. The original specification called for a single axis drift requirement of 1 arc sec/sec (3σ). This was subsequently relaxed to a requirement of 1 arc sec/sec (2σ).

In summary, analyses of the type which have been briefly outlined here resulted in the following pointing system performance specifications (for 0.95 probability):

LOS accuracy = 200 arc sec
 LOS drift = 1 arc sec/sec
 LOS jitter = 0.32 arc sec rms

SYSTEM DESCRIPTION

The Teal Ruby sensor is mounted on the AFP888 spacecraft. These, together with several secondary experiments, comprise the Teal Ruby Space Vehicle System (SVS). The SVS will be deployed at the apogee of an elliptical orbit. Following this, the spacecraft propulsion module will circularize the orbit, which will have a nominal altitude of 380 nautical miles, and inclination of approximately 57 degrees, as currently planned for an ETR launch.

A cutaway view of the sensor is given in Figure 4. For reference purposes, the sensor is approximately 8 feet long by 33 inches in diameter at its widest point, with a gimballed weight of 850 pounds. The major pointing system elements are described next, along with a brief treatment of the sensor itself.

Sensor Assembly

The main sensor elements are the focal plane assembly, the telescope, and the focal plane electronics. These are housed in an evacuated enclosure and cryogenically cooled. The focal plane assembly detects radiant infrared energy collected by the telescope and converts it into electrical intelligence. This intelligence is buffered, multiplexed, and sampled by the focal plane electronics.

The double Cassegrain telescope has an aperture of 20 inches and focal length of 66 inches, yielding an f/number of 3.3. The unobscured field of view is plus or minus 1.125 degrees about the central ray.

The focal plane contains 13 zones, each containing 12 infrared charge coupled device (CCD) chips. All are maintained at approximately 15 degrees K by frozen neon. Each zone is fitted with an optical filter that corrects for flatness and provides spectral filtering. Ribbon cables connect the zones to the focal plane electronics.

The information gathered by the focal plane and processed by the focal plane electronics during a mission is stored and telemetered to ground stations on command, after suitable conditioning. The principal points of interest here are that this assembly and its physical attributes define the pointing system requirements. All accuracy, stability, and servo performance specifications cited in this paper refer to the telescope's central ray, or boresight.

Yoke Assembly

The yoke assembly supports the sensor and provides precision pointing, scanning, and tracking of ground targets independent of spacecraft attitude. The nominal viewing field is from 80 degrees forward to 60 degrees aft of nadir in the intrack direction, and 10 degrees either side of nadir in the crosstrack direction. There are some restrictions to this nominal rectangular field due to spacecraft interference, and a detailed map of usable angular freedom has been developed for planning missions.

Redundant rate gyro assemblies mounted on the yoke structure provide inertial sensing for the gimbal servos in the rate mode, and provide damping signals in the position mode. Torque motors and resolvers provide torque and gimbal angle sensing. These are also redundant, as are the servo controllers.

The resolvers are two-speed, with coarse windings providing 360 degrees of electrical phase shift for 360 degrees of mechanical motion, and fine winding exhibiting 360 degrees of phase for 22.5 degrees of mechanical motion. The signals from these windings are combined and converted by resolver-to-digital (R/D) converters in the servo electronics with a precision of 19.7 arc seconds. The gimbal torque motors are driven by transistor bridges in a proportional pulse width modulated fashion in order to achieve high efficiency. Rated gimbal torque is 215 in-oz intrack, and 206 in-oz crosstrack, which provides single torquer margins of approximately 80 percent in single torquer mode. In the event that additional torque should be required during orbital operations, the second (redundant) torquer in each axis is switched in by command, or automatically under software control if gimbal response becomes too sluggish.

The spindle assembly provides a low friction rotating support attaching the sensor to the spacecraft. The spindle is supported by two duplex preloaded pairs of angular contact ball bearings. Spindle bearings and crosstrack trunnion bearings maintain the precise angular relationship between the sensor and spacecraft, which is required for high accuracy pointing. The bearing and wiring harness constitute the primary parasitic loads for the gimbal torquers.

Gimbal Servos

The gimbal servo loops are closed in the servo processor. Angle and rate commands calculated from the pointing equations in the command processor are compared with measured gimbal angles and rate gyro data. The error signals thus derived are digitally compensated and used to control the torquer driver amplifiers. A simplified block diagram is shown in Figure 5. As shown in the figure, all inputs to the digital controller are sampled, compensated, combined, then converted to a pulse width modulated current that drives the torquer. The resolver outputs are converted to

Error Budget

Consider for a moment what happens during a mission where the sensor stares at and overflies a fixed point on the earth. At the start of scene data acquisition, the stabilization point will be approximately 60 degrees ahead of nadir. The stabilization point will remain fixed on the focal plane as the satellite passes directly over it, and data is acquired until this point is about 40 degrees aft of nadir. During this period, the scene has gone from a condition of substantial foreshortening in the intrack axis, to full-on viewing, to foreshortening as it recedes from view. This means, in effect, that all points above and below the line through the stabilization point appear to move closer, then move away; in other words, the geometry of the encounter introduces focal plane motion of almost all points in the scene. If, in addition, the stabilization point is displaced to one side of the satellite ground track, equivalent geometric distortion occurs across the image. It is only at the instant that the sensor is pointed directly down that there is no geometrically induced motion. Hence, it is at this point that drift due to the sensor pointing system strictly applies.

Error budgets were derived for several different intrack angles, but the one most stressing for the pointing system is for an intrack angle of 0 degrees; i.e., looking straight down. A detailed error budget breakdown is inappropriate here, but Table 1 shows the overall distribution between spacecraft, sensor hardware, and sensor software. The numbers shown are for 0.95 probability, i.e., they are approximately twice the standard deviations. The spacecraft contribution is primarily due to ACDS errors, with contributions from various misalignments. Those listed as servo are due to all hardware and sensor misalignment contributions, while commands refers to errors in spacecraft rate estimation, extrapolation, and errors due to simplifications made in the pointing equation computations for reasons of computational efficiency.

The errors estimates given in the table were provided by Alex Cormack of Rockwell International, who has played a major role in the pointing system development. Portions of the error budget have been substantiated by test, but final verification, of course, awaits flight test data.

Conclusions

The servo design was analyzed and simulated extensively (Ref. 5), and development was carried out using breadboarded flight hardware, both for software development and performance verification, by Rockwell International at Seal Beach, California. Rigid body gain and phase margins are ample, and flexible body margins are predicted to be adequate. Ground testing of the flight system shows essentially deadbeat response, and servo gains can be adjusted on orbit by means of uplink command, should this be necessary.

One aspect of performance verification that is worthy of note is the difficulty of performing adequate test of a system like this on the ground. Due to power and weight limitations, the torque available is severely limited. Before the gimbal servos will operate at all, the large gimbal mass must be balanced almost perfectly. This entails placing the assembly with the intrack axis vertical, and balancing the crosstrack axis with adjustable weights. Low temperature vacuum operation on orbit necessitated the use of special lubricant in the crosstrack bearings, which is subject to oxidation when the bearing is operating. This required the use of a dry nitrogen purge of the crosstrack bearings any time they were subject to movement. Accuracy verification of the spacecraft-mounted sensor could only be done with the spacecraft on its side, which required that the theodolites were operated ten to fifteen feet above ground. Despite these obstacles, the system was thoroughly tested and shown to meet and exceed all performance requirements.

REFERENCES

1. David L. Fried and Richard D. Williams, "Formalism Development and Sample Evaluation of the Mean-Square Clutter Leakage for a HALO-Type Signal Processor," Optical Science Consultants, TR-239 (January 1977).
2. Jerry L. Rapier, "Background Clutter Leakage in a Mosaic Sensor Whose FOV Moves Relative to the Background," Proc. of the Soc. of Photo-Optical Instrumentation Engineers, V. 124, paper number 124-05 (August 1977).
3. W. Jerkovsky, "A Computationally Efficient Pointing Command Law," AIAA Paper No. 33-2208, AIAA Guidance and Control Conference (5-17 August 1983).
4. W. Jerkovsky, "P80-1/TRE Integrated System Analysis (ISA) Modeling Description and Sample Results," TOR-83(3506-21)-6, The Aerospace Corporation, El Segundo, California (27 January 1983).
5. H. Hablani, "Performance Verification of Flexible Teal Ruby Spacecraft with Nonlinear Controllers: an Intermediate Report," SSD-84-0020, Rockwell International, Seal Beach, California (23 February 1984).

ERROR SOURCE	RATE Sec/Sec		POSITION Sec	
	INTRACK	XTRACK	INTRACK	XTRACK
SPACECRAFT	0.0	0.442	65	62
SERVO	0.465	0.386	80	131
COMMANDS	0.182	0.127	17	17
TOTAL	0.499	0.673	104	146
REQUIREMENT	1.0	1.0	200	200
MARGIN	0.866	0.740	171	137

Table 1. 0.95 Probability Error Summary

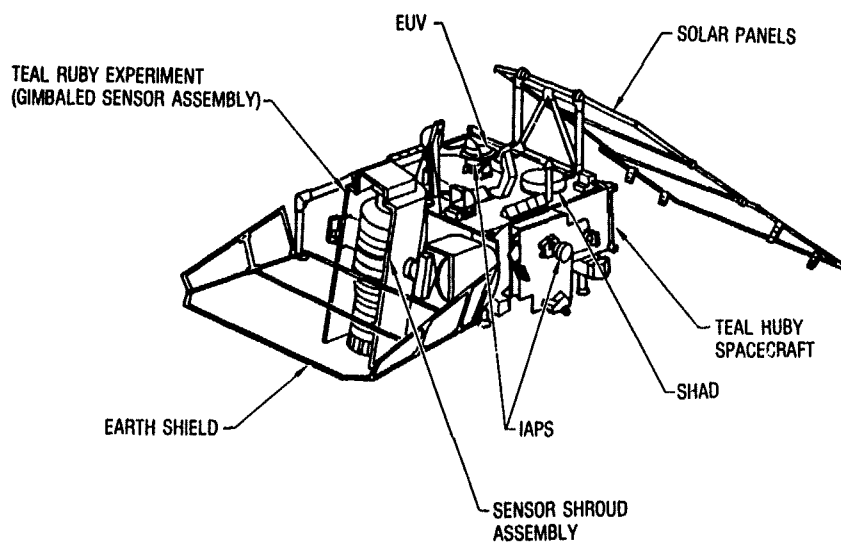


Figure 1. Teal Ruby Satellite

PERPENDICULAR ENCOUNTER
(A → B)

$$K^2 (\sigma_{a1}^2 + \sigma_p^2 + \sigma_{oc}^2) = c^2 \leq \left(\frac{4.9}{2}\right)^2$$

$$K^2 (\sigma_{a1}^2 + \sigma_p^2 + \sigma_{oc}^2) = b^2 \leq \left(\frac{3.7}{2}\right)^2$$

PARALLEL ENCOUNTER
(C → B)

$$K^2 (\sigma_{a1}^2 + \sigma_p^2 + \sigma_{oc}^2) = c^2 \leq \left(\frac{3.7}{2}\right)^2$$

$$K^2 (\sigma_{a1}^2 + \sigma_p^2 + \sigma_{oc}^2) = b^2 \leq \left(\frac{3.7}{2}\right)^2$$

SIMPLIFIED ANALYSIS

$$K^2 (\sigma_a^2 + \sigma_{a1}^2 + \sigma_p^2) = \left(\frac{3.7}{2}\right)^2$$

$$\sigma_{ENCOUNTER} = 1 - \sigma - K^2/2$$

• AIRCRAFT NAVIGATION ERROR

IN-TRACK: $\sigma_{a1} = 0.5$ KM

CROSS-TRACK: $\sigma_{oc} = 0.5$ KM

• EPHEMERIS ERROR, 1-DAY STC PREDICTION

• 1-ORBIT ON-BOARD PREDICTION

IN-TRACK: $\sigma_{a1} = 0.33$ KM

CROSS-TRACK: $\sigma_{oc} = 0.01$ KM

A. ENCOUNTER ERROR ELLIPSES

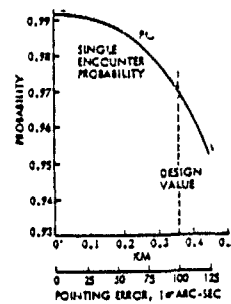
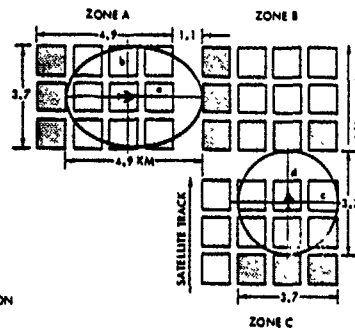


Figure 2. Target Encounter Geometry

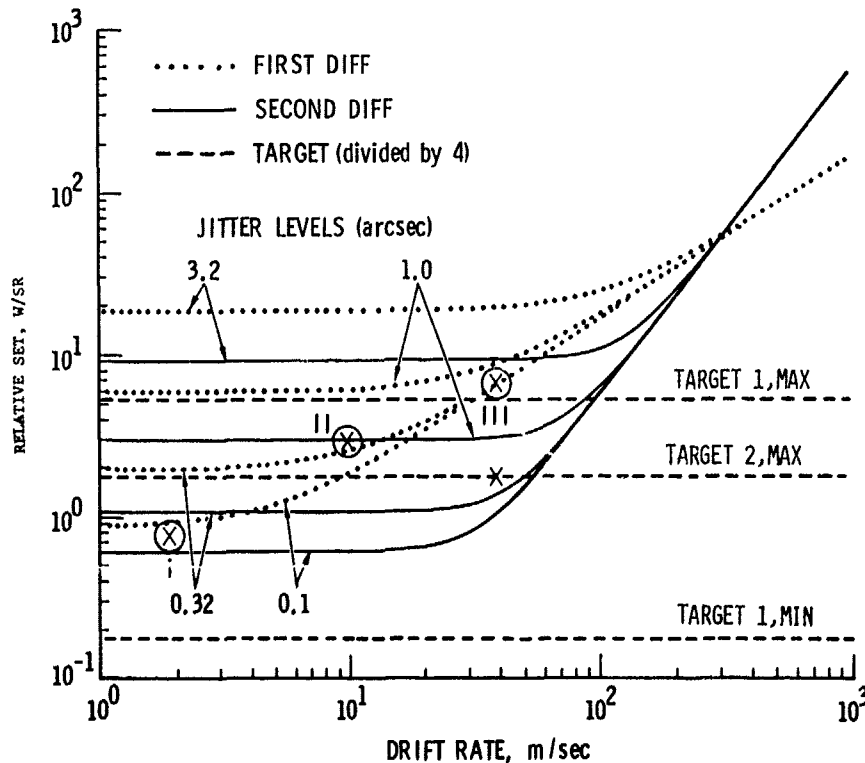


Figure 3. Drift and Jitter Effects

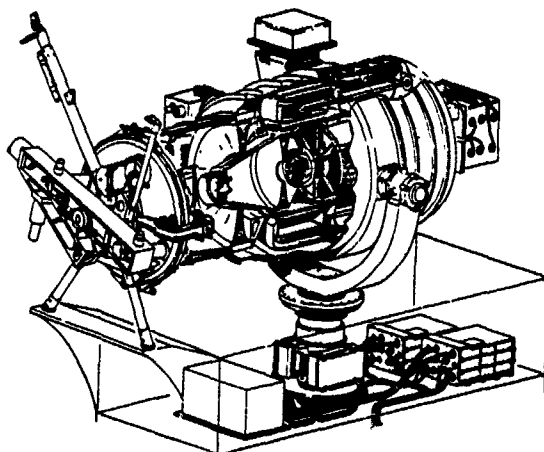


Figure 4. Sensor System Detail

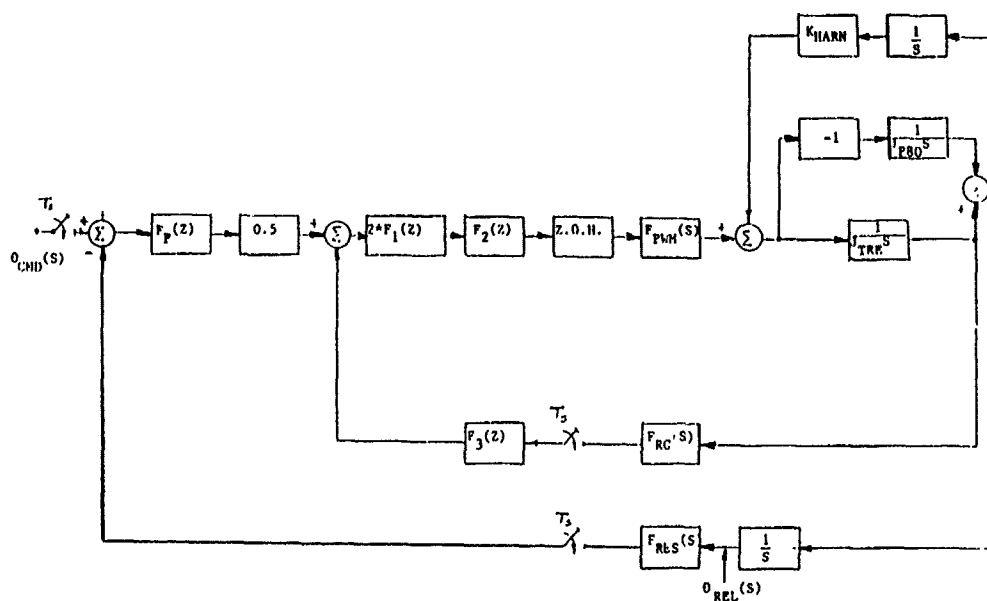


Figure 5. Pointing System - Functional Block Diagram

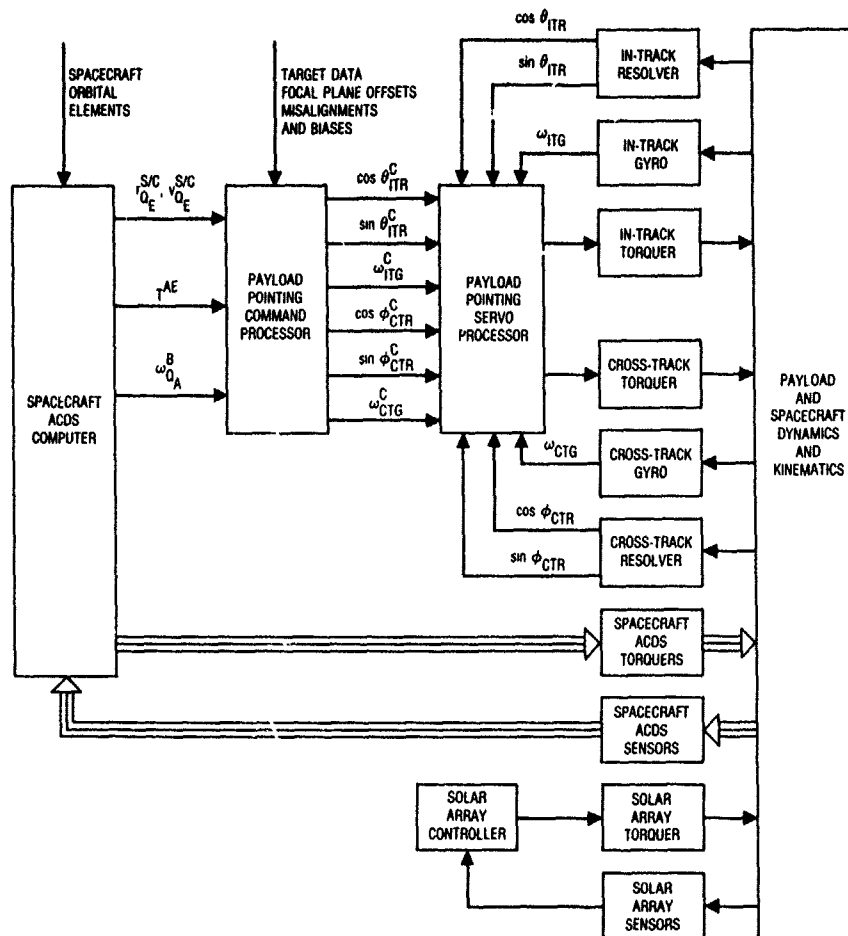


Figure 6. Gimbal Servo Block Diagram

THE POTENTIAL FOR DIGITAL DATABASES IN FLIGHT PLANNING AND FLIGHT AIDING FOR COMBAT AIRCRAFT

by

J. Stone, Marketing Executive
Guidance Systems Division, GEC Avionics
Rochester, Kent ME1 2XX
United Kingdom

INTRODUCTION

The steadily increasing capability of air defence systems in Central Europe is constantly straining the balance between a ground attack pilot's effective mission accomplishment and his survivability. In striving to avoid detection by enemy forces he has been forced into maximising the use of darkness, terrain screening and poor weather - the very conditions which make his task of accurately detecting and attacking targets deep in enemy territory most difficult. This requirement is reinforced by the progressive and demonstrated ability of opposing ground forces to operate during the hours of darkness, as well as the need to make maximum use of expensive resources.

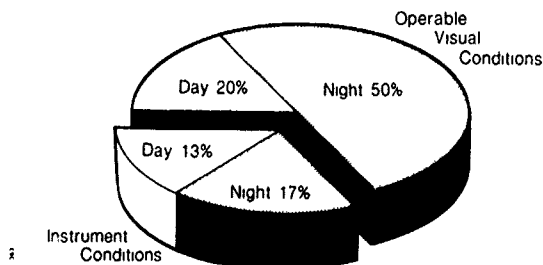


Fig. 1 Average Weather Conditions - Central Europe In Winter

Fig. 1 shows the average weather conditions during the 24 hour period of a mid European winter. Current day VMC attack aircraft could only be effective for some 20% of the time. This operating period could be more than trebled by the use of passive Forward Looking Infra Red (FLIR) sensors. A FLIR image projected onto an advanced wide angle Head-up Display can provide the pilot with a forward view very similar to his normal daylight one, and with the added advantage of a hotspot target detector system. Helmet mounted, image intensifying Night Vision Goggles extend his all-round view, enabling low level manoeuvring flight even in overcast starlight conditions.

These advanced electro-optical night vision systems are relatively simple, passive in their operation and virtually impossible to detect. Nevertheless, there remains some 30% of the time when even aircraft so equipped would be grounded, because of adverse weather conditions.

Full 24 hour all-weather low-level attack capability has been available for some time, demonstrated particularly in the F11 and Tornado aircraft. However, this capability has been achieved at very high cost and hitherto has relied on the use of active, detectable sensors such as ground mapping and terrain following radars.

In order to permit the ground attack the flexibility of full night and weather-independent capability whilst at the same time minimising the probability of detection due to active sensor emissions, GEC Avionics Ltd has for some time been directing its efforts in the overall field of stealth operations and has been innovative in developing the concept of Total Terrain Avionics (T²A).

T²A

T²A is an integrated system in which various operational capabilities, hitherto treated as independent functions, are linked and enhanced by the enormous potential available from an on-board digital database. The range of systems benefitting from this database are outlined at Fig. 2. This capability has devolved as three key technologies have reached sufficient maturity to make their introduction into the airborne environment a practical exercise. These enabling technologies centre on:

- o Digital Mass Memory.
- o Display Capability.
- o Data Processing Techniques.

Enhanced Displays

Map, Radar, FLIR, EScope

Navigation

TRN, TF, TA, GPW

Weapon System

Passive Ranging, SOM Targeting

Defensive Aids

Threat Location

Management

Mission Planning
Mission Debrief
Embedded Training

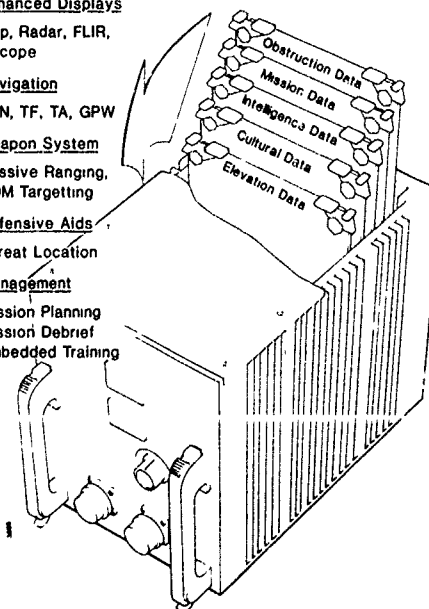


Fig. 2 Database Concept

Digital Mass Memory

The first development is the storage capability of aircraft digital mass memories. Only relatively recently have such memories reached a practical stage where the large amounts of data needed can be stored, rapidly retrieved and processed.

Practical experience and theoretical analysis have led to the choice of solid-state as the data base memory medium. Superior operational and logistic benefits accrue from using semi-conductor memories rather than magnetic tape or laser disc. Because of the widespread and rapidly expanding use of non-volatile semi-conductor memory and its rapid development, a future increase in density and speed coincident with reduction in cost offers considerable benefits in terms of memory expansion and flexibility of use.

The main advantage of video cassette as the memory medium is the large storage provided, around 1G bit, at a relatively low unit price. In addition, if a VHS standard were adopted, logistic advantages would accrue from the multi-sourcing of a cassette hardly bigger than a conventional audio one. However, there are a number of significant disadvantages which make a video cassette reader unattractive in the airborne application.

Environmental performance is limited by the mechanical construction of the head drum assembly in helical scan recorders and the tape material itself. Hermetic sealing of the cassette package might extend the tape's lower operating temperature but high temperature operation would still be severely limited, probably to 71°C. The player would need a heater in order to prevent condensation which would otherwise rapidly degrade the tape by creating drop outs. It is likely that a strict maintenance procedure of cleaning the player would be required in order to prevent the build up of oxide on the heads. The tape transport speeds, unidirectional play characteristic, head drum synchronization time and the serially recorded nature of the data all have an adverse effect on access time. Editing the tape, by overwriting data with new data is a complicated process requiring quite sophisticated equipment and a reasonable level of operator skill.

The laser disc medium has the advantages of large storage capacity (300M bytes/side) and low cost disc reproduction. However there are disadvantages which include hermeticity, temperature range, altitude ceiling and vibration effects.

Hermeticity considerations may require a trade off against altitude which at present is limited to around 4,500 metres. This problem arises from disc construction techniques. Temperature range at present is limited to 10°C to 45°C mainly determined by the focussing arrangements of the laser when reading from the disc.

Vibration problems are associated with read head construction and can be severe. Spot size on the disc is of the order of $\frac{1}{2}$ -2 μ m and the focus and position of the read head must be constrained to limits of this order. The solution might reduce the storage capacity of the disc significantly.

The mechanically complex system needs regular cleaning and calibration, adding to the overall cost of ownership.

These problems will almost certainly be overcome but the timescale for achieving success is uncertain. The best forecasts are 2-3 years for prototype systems with production -2 years later, giving a possible 3-5 year timescale. GEC Avionics are keeping a close watch on the development of these systems and making provision for their inclusion when and if their potential is fulfilled.

EPROM storage has two distinct advantages over the previous options: it operates satisfactorily over a wider temperature range and has a substantially faster access time. By its very nature this type of semi-conductor memory allows any byte to be accessed within the same time period, typically 250ns. The reliability will be superior to a magnetic tape system and the cost of ownership lower. Although semi-conductor non-volatile memories are cycle limited, the number of write cycles available will far exceed any possible requirement. UV EPROMs, when packaged in hybrids, give a memory unit that is of similar size and capacity to a digital recorder but consumes only a third of the operating power. A number of suppliers are available, permitting multiple sourcing. Because EPROM technology has not yet reached maturity, the future looks very encouraging in terms of increasing density and reducing cost. 1M bit military parts are currently flying in GEC Avionics equipment and 4M bit devices are expected to be available during 1988.

The major current limitation associated with the use of UV EPROMs is the time required to reprogramme the full store. Memory manufacturers have programmes aimed at reducing both erasure and programming time which should minimise this limitation.

Electrically Erasable PROM

EEPROM has the same operating advantages as UV EPROM with the addition of faster write times and no requirement for erasure. The complete memory can be re-programmed in a few minutes using the page mode facility linked to a multiplexing system. In this respect it is the fastest option. The read access time is similar to that of EPROMs. The main disadvantage of EEPROM at the moment is its high cost relative to the UV EPROM.

It is worthwhile considering the long term forecasts of certain device manufacturers who are producing both EPROM and EEPROM, together with the individual opinions of members of working parties considering non-volatile memories. In the short term, although the relative cost of EEPROM to EPROM may not change by much, the real cost of both memory types will fall significantly. As a result the cost of manufacturing a hybrid package will become the largest single cost element in the complete memory and the cost difference between EEPROM and EPROM will become less significant. In the longer term, perhaps by 1990, the improvement in EEPROM technology and the increasing demand will make EEPROM similarly priced to EPROM.

Displays Capability

The second development is in the dramatic advance of information display technology using high brightness, high resolution, shadow mask colour CRTs.

Capable of either cursive, raster, or a combined format, complex colour images can be displayed such as low level aeronautical charts with mission overlay symbology, and viewable in all ambient lighting conditions. Such equipment can also show a forward Looking Infra-Red (FLIR) picture to the same resolution as might be achieved on a monochrome display. This multi-function, colour technology permits vast amounts of complex information to be presented in easily discernible, readily assimilated form.

Data Processing Techniques

Thirdly, between data storage and data display, extensive processing is required. The advent of very large scale integration of digital circuitry techniques, combined with the latest high speed digital processors, enables the complex manipulation of vast amounts of data to be accomplished in real time.

The Total Terrain Avionics concept uses the potential available from digital databases in both flight planning and flight aiding. It provides the operational aircrew with an integrated, covert system introducing new or improved capability. The system can be broken down into four general areas.

- o Database.
- o Functions.
- o Displays.
- o Mission Planning.

Database

Fig. 2 shows how the database can be considered as a series of function files of information to meet differing operational requirements.

Elevation Data is required to support the terrain referenced navigation and passive terrain following functions in addition to enhancing the capability of both head-up and head-down displays. The main source of this information is DLMS DTED. This is a US Government system based on a matrix of height posts to a grid spacing of approximately 100m. The data is available at an absolute accuracy of 30m and a resolution of 1m which is more than adequate to support the required functions.

The digital database used in the GEC Avionics equipment is prepared in PIXEL format. There are plans in the USA and in the UK for standard digital databases for cultural data to become available from the early 1990s. Present coverage is far from complete and in the meantime other sources are being used. In all cases the source data is converted to PIXEL format during database preparation.

The cultural database can be obtained from a number of sources including digitized chart, Digital Feature Analysis Data (DFAD) and satellite data. The equipment stores the information in such a way that any combination of these sources can be used without hardware changes.

Digitized chart information is prepared by digitizing either the complete paper map or the original "Feature Plane" negatives or positives. Digitizing in planes gives the ability to declutter at Feature Plane level in flight. It also provides an economical and quick method for obtaining large areas of digitized cultural data.

DFAD has a standard vector format and needs pre-processing on the ground into pixel format for use in the system. DFAD allows full individual feature manipulation.

For areas where no recognised maps exist or where available maps are not easily verifiable, satellite imagery can be used. Some pre-processing is required on the ground to select the correct data planes.

In order to minimise the size and cost of the store required to carry the digital database the data is compressed. GEC Avionics have a continuing programme of research into data compression techniques. The essential features are:

- (a) The data when decompressed is a true representation of the original i.e. the compression - decompression process does not affect image quality.
- (b) The compression technique leads to a "real time" decompression algorithm.
- (c) The compression ratio should be as large as possible commensurate with (a) and (b).

At present the compression ratio obtained with good quality map data generated by "feature plane" scanning is 8:1 and meets (a) and (b).

A further improvement in effective storage density can be obtained by omitting map features which are never used in particular applications. Dispensing with redundant features would simplify the digitisation process and would result in improved compression ratios. The customer would then have the choice of a larger coverage or a smaller, less expensive database memory.

Mission specific data, probably including latest intelligence information and chart amendments is compiled at a local ground station installed at station or squadron level. This additional data is then downloaded to a portable programming unit such as a small cartridge which is then carried out to the aircraft to programme the mission store. The display of this tactical information will be in perfect registration with the aeronautical chart viewed on the multifunction colour display. An example of a digitised chart with mission overlay is at Fig. 3.

FUNCTIONS

Terrain Referenced Navigation (TRN)

TRN provides an accurate, autonomous navigation capability, which is particularly suited to low altitude, stealth missions. The only active sensor is a low powered, downward looking radar altimeter, providing a covert, 24 hour all weather facility.

The concept of terrain referenced navigation has been understood for many years and indeed for some time has been exploited in such systems as cruise missiles. The basic principles of TRN are shown at Fig. 4 in which series of height above-ground samples is used to generate a ground profile. This profile is compared with the on-board terrain elevation data base to find a match and hence a position fix.

Three types of measurement data are required:

- o A series of ground clearance measurements. This is currently achieved by sampling the output of a radar altimeter along a stretch of ground track, or transect. Whilst the horizontal interval between samples is not critical, 100 metres might be considered typical.
- o A baro or baro-inertial height sensor to measure any excursions from horizontal flight.
- o A dead reckoning system to plot the relative horizontal positions of the ground clearance measurements. In practice this tends to be the aircraft IN system. The TRN operates in conjunction with the INS to provide an integrated, accurate and continuous navigation facility.



Fig. 3 Map Display with Overlay

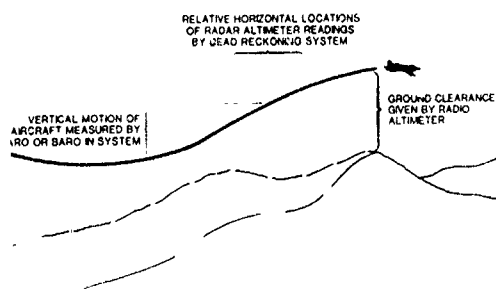


Fig. 4 Terrain Referenced Navigation Principle

The small capacity on-board databases of early generation systems resulted in infrequent fixes over 'patches' of data interspersed with long stretches of inertial guidance. GEC Avionics' SPARTAN algorithm is a third generation mechanism of TRN which allows greater accuracy over these first generation approaches by more frequent fixing, but without resorting to the device of terrain linearisation typical of second generation devices.

Fixes are not independent but use data from preceding fixes to improve the accuracy of the current fix. This also results in the TRN accuracy improving as the flight progresses. Accuracy between fixes, or if fixing is suspended for any reason, is maintained using a Kalman filter. This filter contains a model of the INS error patterns and is used as a predictor of INS performance. Consequently accurate navigation can be maintained during periods

when fixing is not available. During such periods, the navigation accuracy is dependent on three factors: the accuracy of the previous TRN fixing, the amount of previous TRN fixing and the time from the last fix. The filter also produces an indication of the accuracy of the navigation data. This information is used to limit the area of the database searched in producing reference profiles, improving the accuracy of the next fix.

SPARTAN is thus efficient in its use of the information provided by the terrain, but at the same time it is robust. If, for example, a local map error results in an erroneous position update, the error will soon show up as an off-centre peak in the predicted pattern, reinforced from transect to transect. As this becomes apparent the system will correct itself automatically without any need to re-initialise using an acquisition mode. Indeed no special acquisition mode is provided: even from large initial position uncertainties the system will start to generate frequent position corrections as soon as the terrain information permits. Such position uncertainties might result from long flight over water or flat terrain, or when out of radar altimeter range of the ground as is often experienced by the air defence fighter.

The operation of TRN is automatic and generates fixes without operator intervention. Eliminating the navigation task from crew workload allows greater emphasis to be given to other mission accomplishment and survivability aspects. TRN incorporates a steering function enabling the aircraft to be directed accurately to a target or navigation waypoint. The accurate present position derived provides the crew with a real time representation of the local cultural features on the display. TRN performance can be assessed by visually determining the accuracy of the displayed present position and by the ability of the system to direct the aircraft to a designated target or waypoint.

Passive Terrain Following

Accurate knowledge of aircraft position combined with probable future track enables the common terrain elevation working store to be used to calculate a terrain following flight path or Kinematic Flight Path (KFP). This KFP will keep above, but as close as possible to a specified minimum clearance height whilst respecting the aircraft's dynamic limits. Its mechanisation can be the basis for flight director commands for manual terrain following flight or linked into an autopilot for fully automatic passive terrain following.

Immediate operational advantages emerge:

- o Reduction in pilot workload from the demanding task of low level flight increases the time available for other mission management tasks.
- o The need for forward looking, detectable sensors such as TF radar is avoided. Current operational TF systems use a TF radar which typically transmits more than 20 kw at peak power. They make a blanket scan of the terrain ahead within line of sight of the expected flight path of the aircraft and are thus easily detectable. With digital map based TF, only a low power, downward looking radar altimeter is required making detection very difficult.
- o The digital database allows the system to analyse the ground profile "behind the hill" - information not available to current line-of-sight systems. This allows the KFP to take account of re-entrant features with an attendant reduction in exposure to threat defences.

The aircraft track needs to be predicted in advance so that the terrain to be overflown can be accessed from the terrain database. If the aircraft were always flying a pre-planned route with a track hold autopilot mode engaged, the track prediction algorithm would be very simple. However to provide a much more flexible system the GEC Avionics TF system predicts the aircraft track continuously from sensed aircraft dynamic parameters. This allows complete freedom of manoeuvre and departure from planned track whilst still maintaining automatic low level flight.

A possible early problem will be pilot reluctance to trust their safety to a terrain database in the same way that active TF systems were not popular when first introduced. The problem of confidence in the DTED may be overcome by using existing radar TF occasionally to confirm the passive TF. Since the power needed to detect an object increases with range to the fourth power, then the power of an active radar may be dramatically reduced by using DTED to supply long range data. This reduction of power greatly increases stealth. The radar can be switched off totally or pulsed infrequently once confidence in the passive TF has developed.

Ground Proximity Warning

By comparing an aircraft's flight parameters (altitude, velocity, rate of climb/descent, etc.) with its present position, a GPW function will predict when the aircraft will descend below a pilot selected minimum safe height. This can be an invaluable aid in preventing ground and obstacle impacts from such causes as:

- o Late aborts from low level flight in marginal weather conditions.
- o Lack of knowledge of local terrain.
- o Rapid descent to low level - especially through cloud.
- o Temporary distraction whilst flying at low level.

Passive Target Ranging

Knowledge of aircraft precise position from TP4 or GPS, and target bearing from either HUD, helmet mounted sight, or processed FLIR image, real time computations of plan range and height above target can be produced to support the weapon aiming system.

Stand Off Missile Targeting

A DTED subset could be downloaded into a stand off weapon prior to launch enabling the expendable store to use a small memory whilst still retaining the operational flexibility of last minute re-targeting.

DISPLAYS

The ability to manipulate the different types of information contained in the onboard database has led to exciting developments in the way that such information can be displayed to the flight crew. Some of these advanced techniques are described below.

Basic Navigation Display

The basic navigation display provides the pilot with a real time representation of cultural, elevation and aeronautical information such as found on standard aeronautical charts. The display, normally presented in a selectable North or Track orientated form, is based on the aircraft present position and derived through the interrogation and manipulation of the digital database. It is a highly flexible system offering many facilities including:

Scroll and rotation:

Smooth scrolling and rotation of the display without picture break-up during all aircraft speeds and turn rates.

Zoom:

The display can be zoomed to magnify the image.

Scale change:

On demand, the displayed map can be replaced with one of a different scale almost instantaneously.

Look ahead:

The area being displayed can be moved to provide a look ahead facility, so previewing a particular navigation point or target.

Display of present position:

Aircraft present position can be displayed screen centre or offset towards the bottom of the display to provide increased look ahead.

Colour palette modification:

Variable colour palette options allow colours to be manually or automatically selected according to user requirements and ambient lighting conditions. Thus it is possible to match the display colour palette to suit the multi-function display CRT phosphors and filters as well as Night Vision Goggles.

De-clutter:

Where the cultural database has been prepared from separate 'feature planes', a de-clutter facility can be provided editing features from the final display to provide a simplified picture.

Display Enhancements

In addition to the display of basic navigation information derived from the digital store, the fusing of cultural and elevation databases combined with symbol generation can lead to impressive operational enhancements.

Mission and intelligence data can be overlaid in registration by the inclusion of symbol generation. Fig. 3 shows a basic navigation display of 1:500,000 scale chart of the eastern seaboard of the USA with a range of possible overlay facility options. The information can be loaded pre-flight, manually updated and amended, or indeed modified by data link.

The digital terrain elevation data used by the TRN and TF can be used as a source of height data to produce altitude shading on existing maps by assigning colours according to altitude. The boundaries between colours form contours which may be displayed if desired. An example of altitude shading is shown at Fig. 5. The different colour bands correspond to ranges of terrain.

Producing the altitude shading from the DTED has several advantages. The data is more accurate than existing contour shading and it is possible to highlight hazardous terrain by displaying terrain higher than the aircraft in shades of red. In addition, the terrain shading and contours derived from paper map data are no longer required and may be omitted, thus offering significant reductions in data store size.

A passive radar display eliminates the need for undesirable aircraft radar transmissions. For radar bombing attacks the real radar need be switched on

only when the covert pseudo-radar display indicates that the radar could obtain valid target returns. This greatly increases the stealth of the attack as radar emissions need not take place until the final approach to target. The aircraft radar ground returns are simulated by determining which points on the ground are visible to the aircraft. This is performed by simple line of sight calculations based on the mathematics of similar triangles. These calculations determine areas of radar shadow.

A typical passive pseudo-radar display is shown in Fig. 6. The terrain altitude is shaded as in Fig. 5, and areas not detectable are displayed in shades of grey.

Because of the improved low altitude capability of modern surface-to-air systems, it has become essential to take full advantage of any available terrain screening. Real time line-of-sight intervisibility calculations between known radar sites and the operational low-flying aircraft can be used to display threat zones. As an example, Fig. 7 shows the missile sites' capability against an aircraft flying at 70m.

The pilot may easily determine the threats from the display and modify his selected route accordingly either to avoid or indeed attack the threat from the least vulnerable sector. This type of display mode will be invaluable during the mission planning phase if the threats are ascertained prior to the mission. In the event of new threats being located and transmitted to the aircraft during flight via data link, these could be displayed automatically with their associated threat zones. New threats detected en-route could also be noted, recorded for future use or disseminated to other aircraft.

A synthetic image of the terrain ahead can be computed from the elevation data and known aircraft parameters and then displayed perspectively on a wide angle field of view Head Up Display (HUD). This Pilot Low Altitude Terrain Overlay (PLATO) will operate across the full flight envelope of low level fast jet operations.

With precise information this image will accurately overlay the real world as seen through the head-up display or overlay an equivalent FLIR view. It would continue to provide the pilot with terrain information even if he lost sight of the ground temporarily in sporadic cloud or in conditions of poor visibility. Such an overlay can also be most useful in augmenting a FLIR image by picking out ridge lines which can often be confused with other boundaries, such as those occurring between different types of vegetation or when FLIR range is attenuated by meteorological conditions. The resultant composite display can be viewed by the pilot head-up or head-down.

A typical PLATO display superimposed on the HUD is shown at Fig. 8.

Whilst these developments in information display are geared primarily to the demanding and hostile world of low level operations, they can also be used to assist the air defence pilot operating from medium level.

A plan view of the terrain showing sun or synthetic radar shadowing allows the pilot to assess such variables as:

- o The optimum Combat Air Patrol height to fly to minimise terrain obscuration.
- o Areas where the look-down/shoot down missile might lose target tracking after launch due to terrain screening.

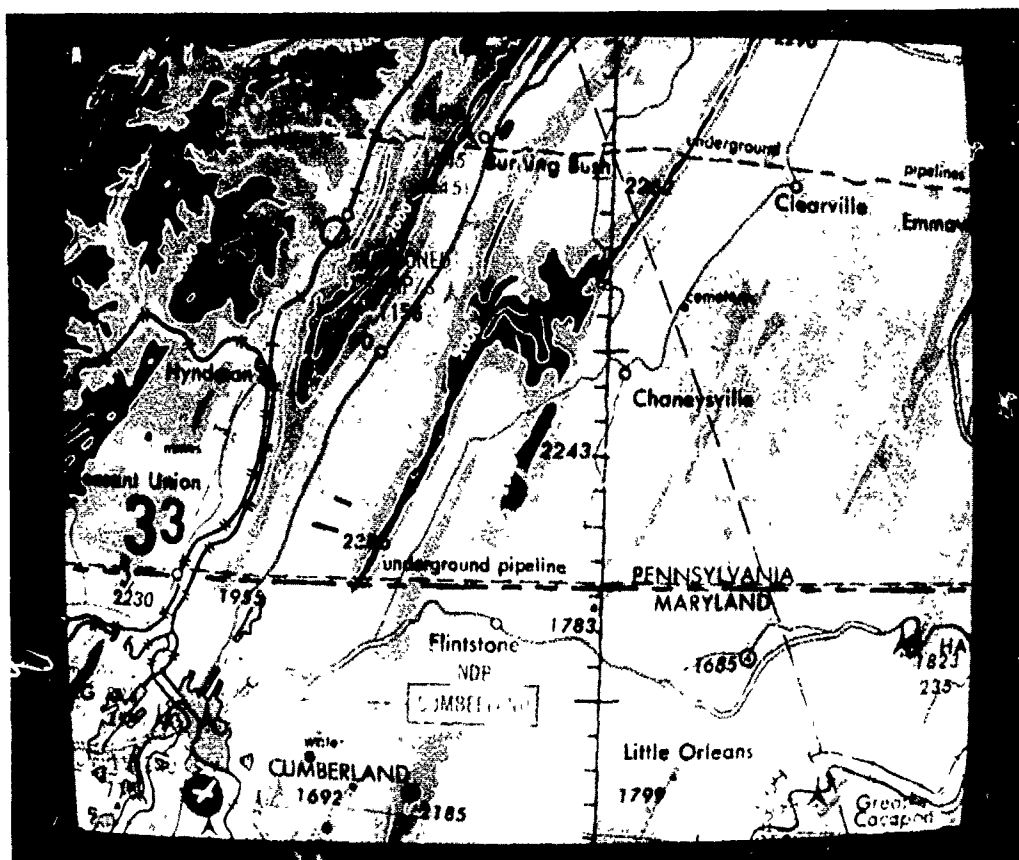


Fig. 5 Altitude Shading Using Digital Terrain Elevation Data

A long range oblique view of an area of terrain can be displayed head-down to present the overall ground profile situation. This three dimensional image is well suited to the display of tactical air defence information such as might be received by military data link.

Mission Planning

The capabilities described above are either under development or in production. The Digital Colour Map Unit (DCMU) is in production for both air and ground applications it is a major step forward in the presentation and management of topographical map and other related information in an efficient, economic and flexible manner. As with the TRN/TF equipments, this is the aircraft/end-user element of a three part system which ensures that a high degree of operational flexibility is retained whilst achieving reliability and maintainability.

Main Ground Station

The three-stage concept assumes that the main topographical database would be prepared off-line and distributed by a central agency who would be responsible for its accuracy, integrity and maintenance. The procedures would be comparable with those currently used for distribution and maintenance of paper maps and of film strips prepared from them. A digital database is simple to maintain and update but mandatory procedures and rules would be needed to ensure that equipment manufacturers or users did not impair the integrity of the database by conversion, manipulation, compression/decompression or other techniques. The responsible authority could exercise control by allowing only approved techniques to be applied in modifying and handling the data. Updates and amendments would be generated by the agency who would ensure that all the users were at the same issue state. Distribution could be tape, disc or over telephone lines with safeguards to avoid corruption of data when being incorporated in the user database. For areas not covered by the Agency database other methods such as digitizing paper maps or their feature planes could be used, subject to approval by the Agency.

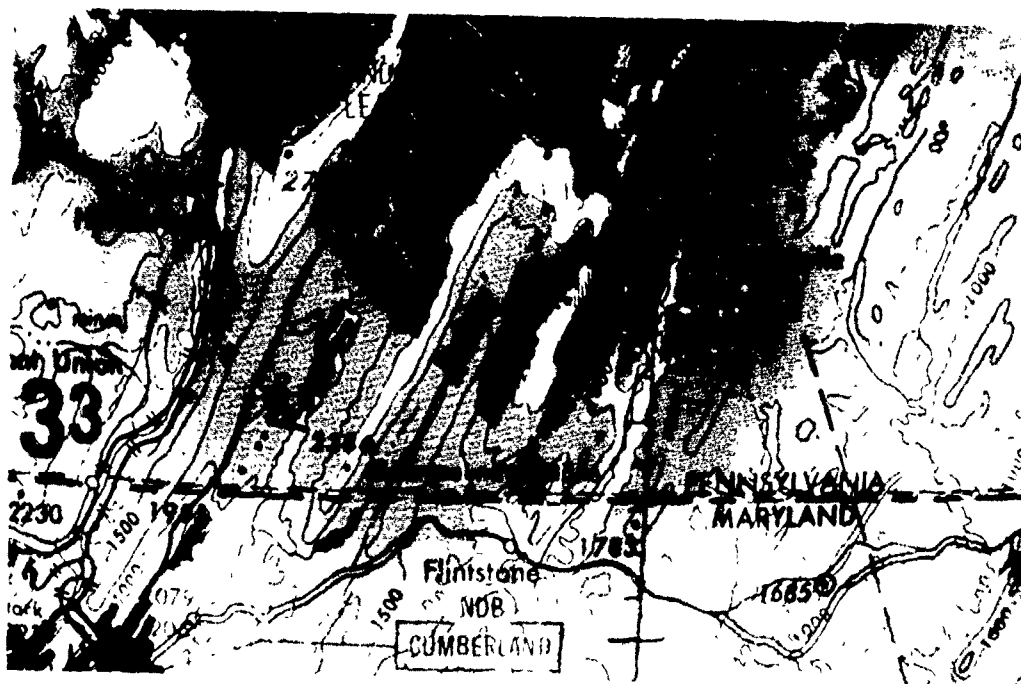


Fig. 6 Pseudo Radar For Covert Operations Using Digital Terrain Elevation Data



Fig. 7 Surface To Air Threat Display

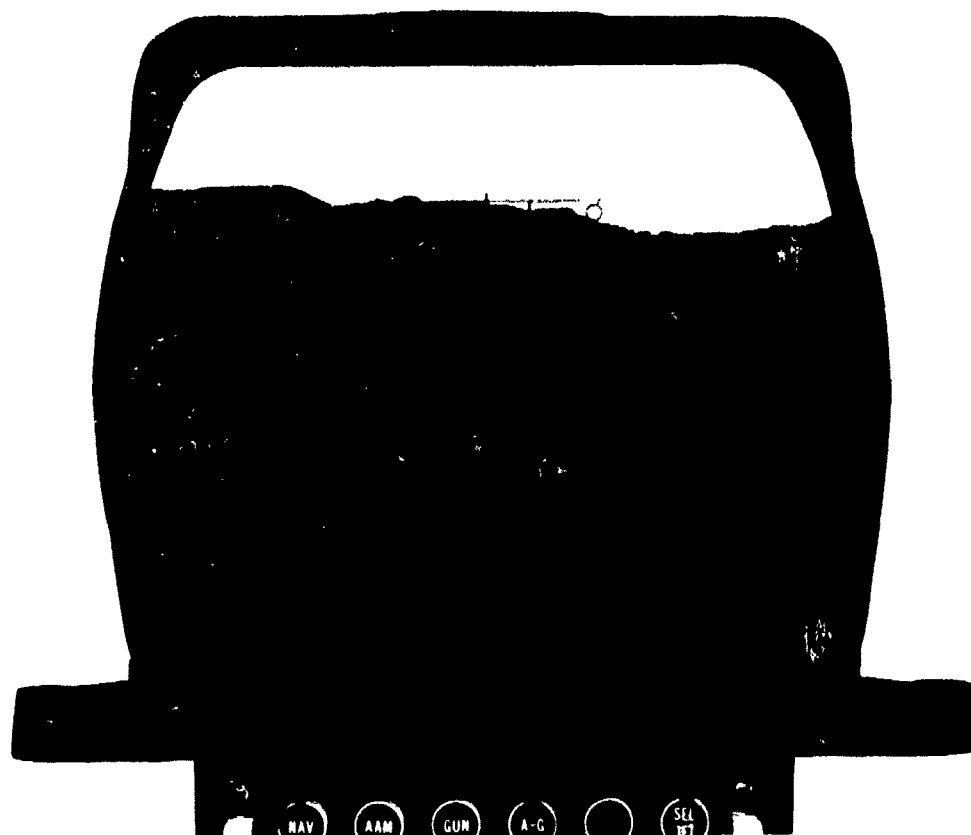


Fig. 8 PLATO Overlay On Head Up Display

Local Ground Station

The second stage is at squadron or base level. A subset of the total agency database covering the full operational area for the unit would be maintained. Facilities would be available to overlay the topographical database with useful information such as intelligence, threat zones, allied and enemy force dispositions and mission-specific information including routes, initial points, targets, alternates and timing points. The overlay would be prepared using a large screen monitor showing the operational area at large map scale and a micro-computer keyboard and cursor to position and enter information (see Fig. 9). Intelligence and threat information could be kept up to date as new information was received and mission-specific information added immediately prior to going out to the aircraft.

Incorporation of elevation data into the system would permit 3-D planning such as perspective displays of the attack phase of the mission to determine the optimum direction of approach. Techniques such as automatic preferential routing, taking account of threats, terrain screening and ground obscuration can be incorporated to suit the operator's needs.

This mission-specific data could be loaded into the aircraft database by means of a small cartridge data transfer module direct, or via the aircraft data bus.

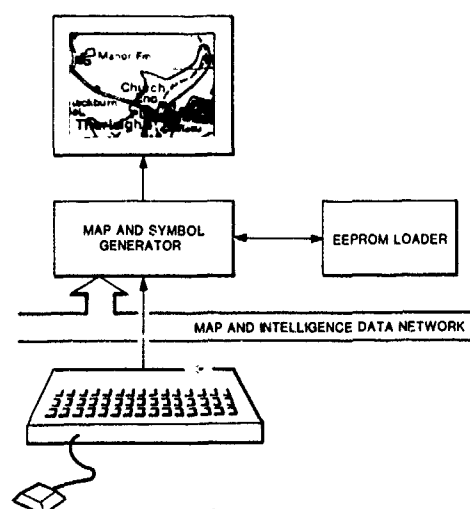


Fig. 9 Mission Planning Station

CONCLUSION

The forces opposing the NATO alliance are constantly improving their capability in the detection, location and interception of attacking aircraft. The ground attack pilot's survival depends upon his ability to conceal his approach by minimising his aircraft signature in all visual, thermal, acoustic and electronic aspects. In attempting to achieve this, he is forced to fly at altitudes and periods of the day which make the very essence of his mission most difficult to accomplish. From this background has emerged the general concept of 'stealth'. The introduction of passive electro optical sensors in the form of FLIR and NVGs has been a significant step forward in aiding the combat pilot in this hostile environment. TF radars enable low level flight in all weathers. However, the FLIR/NVG combination is not all weather and TF radars are not stealthy.

Recent developments in data storage, advanced processing techniques and highly efficient display presentation have been instrumental in enabling covert operations to take place in all weathers. Precise autonomous navigation and terrain following, with the threat of detection minimised, is now available to the modern combat pilot, freeing him to concentrate on successfully achieving the aims of his mission.

The potential for digital databases in Flight Planning and Flight Aiding for Combat Aircraft is recognised and equipments are now being procured for military use. The capabilities described in this paper are all feasible, and are being integrated into GEC Avionics' system of Total Terrain Avionics (T²A).

THE ASSESSMENT AND SELECTION OF INERTIAL SYSTEMS FOR ARTILLERY

Dr Y K Ameen
Applied Science Analytics Ltd
11 Hazlitt Drive
Maidstone Kent
ME16 0EG, UK

G B Symonds

1225 McBride Avenue
Little Falls, NJ
07424, USA

SUMMARY

The paper describes a unified approach to the assessment and selection of inertial system solutions for artillery applications. Whilst the main emphasis of the presentation lies in the field of self-propelled guns (SPG) and mobile observation post vehicles (OPV), it is advocated that similar cohesive treatments of system analysis can and should be applied to other component variants of the artillery system such as rocket launchers, target locating radars and remotely-piloted vehicles. A brief review of the current gun deployment method is presented, followed by a description of the analytical harmonization process carried out to define the inertial system performance requirements. The paper highlights the environmental and mechanical factors associated with the installation of an inertial system on the gun trunnion and the special techniques to be applied in their testing and measurement. The use of the Global Positioning System (GPS) in both stand-alone and integrated modes is briefly discussed.

INTRODUCTION

The function of an artillery weapon system is to fire a projectile from one point to another such that it falls with a prescribed accuracy, consistency and precision, all of which are influenced by four main sources of error, viz:

- a. variations in internal ballistic and gun dynamic parameters.
- b. deviations from the predicted ballistic flight path.
- c. uncertainties in the location of the target with respect to the weapon delivery means.
- d. deviations in gun orientation angles from the Command values

The first two error sources are governed by factors outside the scope of this paper, although their practical magnitudes have been taken into account in order to achieve a well balanced perspective on the third and fourth and on the optimum overall error budget. In contrast, from technological and operational viewpoints, the location of target and weapon delivery means and the static lay of the gun barrel are inherently concerned with positioning and orientation determining processes. Current developments in both the SPG and the OPV indicate that on-board inertial systems will greatly enhance the operational value and survivability of these elements in the future battlefield environment and that careful harmonization of their respective requirements may lead to benefits in terms of overall system performance and cost effectiveness.

DEPLOYMENT

Conventional gun survey

Guns in the UK Royal Artillery (RA) are surveyed¹ using the Position and Azimuth Determining System (PADS), a full inertial navigation system (Figure 1). The PADS is driven from a Survey Control Point to a chosen battery position, a director is set up behind the vehicle and a north reference is obtained via the mirror on the outside of the unit; a reference object (RO) is chosen and surveyed in from the director and the grid position is obtained from the PADS display. When the guns arrive on the position (Figure 2), a battery director, having received its reference from the PADS director, is used to pass line to each gun. This is done by aligning the Indirect Fire Sight Periscope of each gun in turn to the director and by sending a runner to shout to each crew the relevant bearing angle. Finally, each gun surveys its own ROs using the periscope.

Disadvantages

There are clearly several disadvantages with this method of deployment. First of all, the guns are vulnerable as they are constrained, by the availability of the survey party, to close deployment in batteries of eight or, at best, sections of four. As a result they are easier to detect, present a more worthwhile target to engage and are likely to sustain more damage than guns deployed in dispersed positions. The guns are, of course, essentially out of action whilst on the move between battery positions but their availability and response are more specifically limited by the time required for survey and for the determination of their respective positions from the battery centre (at best 10 minutes); should the battery be deployed in well concealed positions such as towns, line-of-sight problems will require several line passes and this will extend the time significantly. In the event of the guns being exposed to counter battery fire, the situation may potentially be worsened as a result of damage to the sight optics or of obscuration/destruction of the ROs. Although PADS is comparatively precise in its position and azimuth determination the benefit of accuracy is prejudiced by the errors incurred when passing line in the manner outlined. Furthermore, modern SPGs are required to fire at high burst rates, typically 3 rounds in 10 seconds, and conventional sight referenced laying systems do not readily permit the gun layer to achieve this with the required level of accuracy and consistency. It should be noted that survey and on-board

laying will often present ergonomic problems. Personnel will be required to operate precise optically-referenced equipment in the dark, under adverse weather conditions, wearing Nuclear, Biological and Chemical (NBC) protective clothing and/or Arctic weather equipment. This will further impact on response and accuracy as defined above.

Target Observation

One of the major methods of current target observation relies on a Forward Observer (FO) in a dug-in position using target surveillance equipment. The observation position and the target bearing from it are referenced to a PADS mounted in a nearby vehicle whilst the range to the target is determined using a portable laser range finder. This approach to target observation has a number of disadvantages including slowness of response, inflexibility of operation, difficulties of passing line and vulnerability of the FO.

Flexible Operation

Current developments in the surveillance and target acquisition area are directed towards the use of highly mobile OPVs from which targets can be acquired and reported with an accuracy and response commensurate with modern Command, Control and Communication Systems. These OPVs will operate in a fluid tactical manner, using sophisticated observation and ranging equipment integrated with on-board position and orientation sensors.

Operational analyses have served to highlight the limitations of the current gun deployment methods and have led to a comprehensive examination of a concept whereby SPGs deploy in pairs, operate in a dispersed and more flexible manner and are able to come into and out of action within seconds, thereby avoiding detection and counter battery measures. To achieve this, each SPG would require a system known as the Automatic Gun Laying System (AGLS) comprising an inertial navigation system mounted, in an integral fashion, with the elevating mass. When viewed in complement to the need for a similar facility on the OPV it can be seen that modern inertial technology is destined to play a critical and major role in the future deployment of such artillery systems.

SYSTEM ERROR ANALYSIS

Bias and Random Errors

In discussing errors within the artillery system, it should be noted that traditional accuracy requirements have been primarily concerned with the random errors in the system. For this reason orientation and position accuracies have generally been stated in terms of Probable Error (PE) and Circular Error Probability (CEP) respectively. These measures of accuracy, however, fail to capture the bias errors in the system and, although a statement about the relevant mean values is a step in the right direction, this alone may prove inadequate in practice since it does not permit a complete trade-off between the bias and random constituents of the total allowable error. This problem is overcome by defining both of the following for each of the sub-system errors of interest:

- a. rms value for the error.
- b. percentage of errors not to exceed a certain value (eg 99% limit)

In the following analyses all values quoted are rms

General Composition of Weapon Delivery Errors

The weapon delivery accuracy requirement must take account of factors such as tactical procedures, target details, the type and quality of ammunition to be expended, the target range and the damage level to be inflicted. Since numerous error sources contribute towards the overall weapon delivery accuracy a balanced apportionment of the error budget must be made to avoid a serious mismatch in emphasis between the different components of the system. In particular, the definition of the performance characteristics of the inertial systems for the SPG and the OPV must be based on a joint consideration of the influences of these on overall delivery effectiveness. As an illustration, consider a situation where a projectile is required to fall within a distance X from the target position. This error is defined by the expression:

$$X = \sqrt{\sigma_1^2 + \sigma_2^2 + \sigma_3^2}$$

where σ_1 is the gun delivery error,

σ_2 is the gun location error,

σ_3 is the target location error.

The gun contribution, σ_1 , is composed of a number of major error sources, so that:

$$\sigma_1 = \sqrt{\sigma_{11}^2 + \sigma_{12}^2 + \sigma_{13}^2 + \sigma_{14}^2 + \sigma_{15}^2 + \sigma_{16}^2}$$

where σ_{11} , σ_{12} are errors due to uncertainties in the angle of departure and velocity of the projectile respectively at shot exit,

σ_{13} , σ_{14} are errors due to assumed air temperature and density respectively.

σ_{15} , σ_{16} are errors due to assumed wind velocity and projectile weight respectively.

Table 1 provides an appreciation of the sensitivity of some of these error sources to the ballistic trajectory computation for a 155mm SPG firing to a range of 15km.

Error Source	Error Magnitude	Delivery Error, m
Wind Velocity	1 knot	20
Air Temperature	1%	10
Air Density	0.1%	8
Muzzle Velocity	1 m/s	25
Projectile Weight	0.5 kg	20

Table 1. Delivery Error Sensitivity (155mm Howitzer firing to a range of 15km)

Harmonization of Inertial System Performance

The process of harmonization is designed to ensure compatibility and optimization of the SPG and OPV inertial system solutions and to take account of the logistic and economic factors surrounding the need for the avoidance of system proliferation.

As a starting point, and on the basis of factors such as ammunition lethality and re-supply, it is considered that a typical mission would require a weapon delivery accuracy of 75m at 15km range. It is further noted that the gun and target location errors exhibit equal sensitivity to the overall error and that, qualitatively, an excessive relaxation in gun delivery error would either unduly restrict the mobility and flexibility of operation of one or both weapon elements or demand an intolerably high cost solution.

At the SPG, despite the move towards on-board ballistic sensors and processing, the limitation of these, together with remaining uncertainties in meteorological data, would result in a residual of approximately 50m (excluding that due to angle of departure uncertainty) at 15km range. The angle of departure uncertainty, α , is given by:

$$\alpha = \sqrt{a_1^2 + a_2^2 + a_3^2}$$

where a_1 is the absolute error of the inertial system,

a_2 is the characteristic error due to gun 'jump' and barrel flexure,

a_3 is the combined error due to trunnion axis orthogonality and control system errors.

A rigorous quantification of the dynamic errors contained within a_2 is not yet possible, experience so far gathered, suggests that its contribution is close to 2 mil (6400 mils = 360 degrees). Trials have confirmed that the errors constituting a_3 can be limited to 0.5 mil each. If the practical upper limit of azimuth accuracy at the inertial system interface is assumed to be 1 mil then it can be seen that the error in angle of departure is approximately 2.3 mil, resulting in a σ_{11} of 35m and a gun delivery error, σ_1 , of 61m.

At the OPV it can be seen that the target location error (TLE), σ_3 , is given by:

$$\sigma_3 = \sqrt{\sigma_{31}^2 + \sigma_{32}^2 + \sigma_{33}^2}$$

where σ_{31} is the position error of the OPV,

σ_{32} is the error due to inertial system heading error,

σ_{33} is the error to the OPV observation sensor errors.

Since the target is normally observed from a distance of less than 5km an azimuth error of 1 mil has little influence on target location accuracy. Similarly, errors associated with the observation sensors are comparatively small and it can be deduced that the target location accuracy is dominated by the navigation performance of the inertial system.

Following on from the gun delivery error analysis, and recognising the equal sensitivity of gun and target location to weapon delivery accuracy, the 75m requirement suggests a 30m position error apportionment to both σ_1 and σ_3 . Since the flexible operation concept requires similar tactical mobility for both elements, in terms of distances and durations of moves, it can be concluded that the same quality of inertial system would be needed on both vehicles. For practical scenarios this leads to a performance requirement of between 0.25-0.30% of distance travelled, commensurate with the proposed azimuth performance required at the gun.

To illustrate the cost effectiveness of adopting this performance standard, based on the 75m requirement, and to further examine the impact of a relaxation in OPV performance, in view of its comparative insensitivity to heading accuracy, a sensitivity analysis was carried out. Its aim was to determine the effectiveness of artillery fire in terms of ammunition expenditure and missions accomplished

for varying TLE. The value of the study is illustrated in the following example where two typical well-defined targets were considered. The targets were:

- a. an Infantry Company (350m x 50m) - Mission A
- b. a Regimental Artillery Group (100m x 100m) - Mission B

A successful mission was defined as the engagement of a target with unadjusted fire at a range of 17km from the gun to achieve a 40% casualty level. Fig 3 shows the relationship between ammunition expenditure and TLE to achieve this success level. As expected, the trends for both mission profiles show that the number of rounds increases with TLE, the effect becoming highly significant beyond errors of 50-60m. A good illustration of this is the comparison between a TLE of 30m and 120m during Mission A where twice the ammunition would be expended to achieve the same effect. The consequent relationship between the number of successful missions and TLE for a given availability of ammunition at the SPG shows a factor of 3 improvement between the same two scenarios.

In a short period of Operational use the extra ammunition cost would soon outweigh the cost differential involved in the selection of a lower performance inertial system for the OPV. Furthermore, the increased demands placed on the ammunition logistic train and the timely engagement of subsequent targets would soon become critical factors.

In summary the recommended performance level for the SPG and OPV can be defined as:

- a. Initial gyrocompassing accuracy of 1 mil.
- b. In-run azimuth error growth of less than 1 mil/hr.
- c. Navigation accuracy of 0.3% of distance travelled.

SYSTEM ASPECTS

Categories

The inertial system categories currently available for land navigation and orientation applications can be broadly classified as:

- a. Heading Reference Systems (HRS)
- b. Attitude and Heading Reference Systems (AHRS)
- c. Inertial Navigation Systems (INS)

The use of odometer data is a common feature of these systems, either as the sole source of incremental displacement information or as a closely integrated complement to the accelerometer derived data of a full INS. In the latter case a zero velocity update (ZUPT) facility is often available. Each of the categories employ both gimballed and strapdown mechanisations based on a range of gyro technologies. The more dated technology systems employ a distributed approach, separating the directional and north seeking gyros into separate modules. Since this particular sub-category, together with the HRS, present general limitations in terms of performance, ergonomics and reliability they have been excluded from consideration in the following discussion.

In contrast, the achievable performance from both AHRS and INS lead to their firm consideration for the artillery application and these will now be considered in more depth. Since the strapdown AHRS can generally be treated as a sub-set of the strapdown INS, for a given quality of sensors, and since the current and projected trends in the gimballed INS show that acquisition and ownership costs will be prohibitive, it is considered that a more poignant comparison is that between gimballed AHRS and strapdown INS.

Gimballed AHRS

Most currently available AHRS rely for good performance on a single high quality gyro used in a dual role, viz for gyrocompassing and as a directional gyro. When the system is switched into the alignment mode, the input axis of the gyro is constrained to lie in the horizontal plane for the gyrocompassing process. Upon entering the navigation mode of operation, this axis is re-oriented to be perpendicular to the horizontal plane in order to measure the heading changes following the initial alignment.

The locally level reference is established using the outputs of null-seeking electrolytic bubble levels or low grade accelerometers for positioning the roll and pitch gimbals. This reference is subsequently maintained using gyroscopic stabilisation in both pitch and roll axes, the gyros used being of a much lower quality than the azimuth gyro. Synchro or resolver outputs provide heading and elevation information. Position is calculated using a recursive software algorithm which sums the resolved incremental position, using an odometer input, to provide northings, eastings and altitude.

Because of this simple approach to the navigation function, the AHRS error propagation is dominated by only four major error sources as follows:

- a. Odometer scale factor (OSF) error (k)
- b. Crab angle errors in azimuth (δ) and elevation (θ)
- c. In-run azimuth drift
- d. Initial heading error

Strapdown INS

The error propagation within the strapdown INS is more complex than that observed for the AHRS for two main reasons. Firstly, the INS is Schuler-tuned so that most errors propagate in an oscillatory manner and, secondly, the mixing of the odometer and inertial data is performed using a statistical estimation process in the form of a Kalman filter. In addition, there may be significant coupling between different error sources, a factor not normally associated with the simpler AHRS mechanisation. Because of these differences, the strapdown INS deals with the uncertainties in the OSF and the crab angles in a manner totally different from that of the AHRS. The system structure also modifies the effects of the initial heading error and the azimuth drift so that they are not as easily predictable as is the case for the AHRS.

Calibration of OSF/Crab Angles

In order to carry out the navigation task, both AHRS and INS require initial nominal values of the OSF and the crab angles. It is in the subsequent use of these values that some of the fundamental differences between the systems become readily apparent. When used in the field, the AHRS requires a special periodic calibration run to refine the nominal or stored values of the OSF and the crab angles. This usually takes the form of a single or return journey between two known points separated by a distance of a few kilometres. The uncertainties in the OSF (k) and the azimuth crab angle (δ) are then determined.

It should be noted that such an approach makes no allowance for the effects of the initial alignment error and the in-run azimuth drift so that significant errors could be present in the calculated values of k and δ . Even in the absence of such errors, the calibration technique is highly limited in establishing the correct values of these parameters. If the calibration runs are repeated many times over the same terrain, widely different estimates of k and δ can be generated. Figure 4 shows the distribution of such estimates for the OSF uncertainty which can be observed to contain both bias and random components. In addition, dramatic changes in bias and random variations in the measurement of the OSF can arise when calibration runs are repeated over different types of terrain (Figure 5) and surface conditions (eg wet, dry, etc. - Figure 6). It is for example possible that the bias uncertainty in the measurements performed over a specific route may be both large and of opposite signs for wet and dry surface conditions. Thus, not only will the apriori calibration based on a single run be unable to measure the OSF without significant bias and random errors but these errors may themselves fluctuate wildly for different surfaces and terrain features. Periodic apriori calibration of the OSF is unlikely to ameliorate the situation appreciably.

In summary, the OSF calibration technique used for the AHRS is unlikely to cope satisfactorily with factors such as:

- (i) Changes in terrain types
- (ii) Changes in terrain surface conditions
- (iii) Wear of tracks
- (iv) Variations in track tension
- (v) Track slippage

The above remarks are also applicable to the estimation of the azimuth and elevation crab angles but for different underlying factors. Here also both bias and random uncertainties due to the initial heading error, in-run drift, load variations, etc., during the calibration run will be present in the apriori measurement based on a single run.

In contrast to the AHRS, the INS calibrates the OSF and the crab angles in real time using the Kalman filter. As a result no gross errors, such as large bias offsets in the calibrated values, are likely to be present and, providing the calibration parameters are modelled satisfactorily, the effectiveness of the calibration process is limited only by the inherent time constants of the filter. Transient errors in the calibration parameters may arise following abrupt and large changes in the terrain characteristics but the resulting navigation errors are not expected to be significant under most conditions.

Initial Heading Error/Azimuth Drift

The AHRS mechanisation offers little scope for estimating the system heading error during navigation. The azimuth drift is usually different for mobile and stationary conditions, being larger when the vehicle is moving. Thus, even if special vehicle stops are allowed for azimuth drift calibration, measurements thus taken are unlikely to improve the navigation performance appreciably.

For the INS, the problem is overcome automatically in real-time using the Kalman filter wherein a mixture of inertial and odometer data is used to estimate the relevant system errors during the journey. Although ZUPTS are not a mandatory requirement, they can be used on an opportunist basis to enhance performance and to provide a powerful reversionary capability in the case of an odometer failure.

Azimuth Scale Factor Uncertainty

One area in which the gimbaled AHRS could have significant performance advantage over many strapdown INS is in relation to the effect of the azimuth gyro scale factor uncertainty. For most gimbaled AHRS, this error is either negligible or of low magnitude; in the case of the strapdown INS, this is true only for systems employing sensors of the quality of the ring laser gyro. For systems using mechanical gyros, the error could be significant depending on the vehicle mission profiles, the azimuth scale factor uncertainty not normally being modelled within the Kalman filter.

Ring Laser Gyro (RLG) Systems

A factor which is of paramount importance for the AGLS is the continuous availability of the inertial system during the entire period of operation. System reliability is therefore a foremost issue in the selection of inertial equipment for the AGLS or similar applications.

The subjects of system availability and reliability are fraught with difficulties of definition and subjective interpretation. On intuitive grounds alone, however, there can be little doubt that electro-mechanical devices must be inherently less reliable than those employing 'solid-state' technology and that, for comparable sensor technologies, strapdown mechanisms offer the greater potential for reliability than gimbaled.

Of all the sensors suitable for the strapdown INS, the RLG offers the maximum promise in terms of cost, performance and reliability for the particular land applications under discussion here. Indeed, the potential for low acquisition and ownership cost combined with high performance has been the main driving factor behind the development of the RLG world-wide. The RLG based strapdown INS is essentially a 'fit-and-forget' system requiring no routine calibration or maintenance. It has, in principle, only one mode of operation after the completion of initial alignment whose duration can be tailored to suit prevailing operational constraints. The system has an inherent capability to recover from interrupted or truncated alignments and, since it is capable of exploiting opportunist vehicle stops, which are mostly inevitable for land operations, mandatory ZUPTs are not necessary. The system can potentially maintain high levels of orientation and navigation performance over long periods of time thus providing a greater degree of operational availability to the field commander.

Environmental Considerations

Although inertial systems cover a broad spectrum of applications, there is one area in which the gun environment is fundamentally different from virtually all others. This concerns the use of a trunnion-mounted inertial system to provide high orientation accuracy when subjected to severe and repeated gun firing shocks, interspersed with long periods of travelling vibration. Operation under burst-fire conditions (e.g. 3 rounds in 10 seconds) also introduces additional performance requirements. It may be noted here that, in times of conflict, the ammunition expenditure per gun, according to some forecasts, may be in excess of 400-500 rounds in one battlefield day alone. Even in peace-time, strict safety must be observed and a large number of rounds may be fired by a gun in training, demonstration, exercises, testing, etc.

Extensive investigations have been carried out to assess the effect of the gun shock and vibration on the performance of an inertial system. Instrumented firing trials have yielded useful data about the nature and effects of the firing shocks. It was, for example, observed in one instance that a sensor specification based on a shock magnitude-duration criterion (eg 200 g, 50 ms) proved inadequate to cater for lower magnitude shock levels at higher frequencies.

Amongst the many time and frequency domain techniques available for shock data analysis, maximum preference has been given to the approach based on spectral density evaluation. This technique has been found to be particularly useful in establishing the natural modes of the trunnion structure over which the inertial system is required to be mounted. Figure 7 represents the autospectrum of the shock transient experienced at the inertial system interface to the mounting bracket as a result of a 155mm SPG firing a high strength charge. Significant resonant peaks can be observed in the frequency range 100 Hz to 4 kHz, indicating appreciable presence of high shock levels of very short durations.

Although the vibration environments of the SPG and the OPV are expected to be similar in character, there is one noticeable difference in their dynamics under mobile conditions. For the OPV, movement of the turret and the mantlet whilst on the move is a special requirement. The system performance in this case will be more sensitive to its location in the vehicle (ie mantlet, turret or hull) than for the SPG where these motions are inhibited during travelling.

An important consideration, this time in common with many land applications, is that of system operation over sustained periods and with a large number of switch-ons throughout its lifetime.

An operational constraint, of particular importance to the strapdown system, is the need to maintain the specified system gyrocompassing accuracy when the vehicle is positioned at upto $\pm 10^\circ$ in pitch and roll. The angular rate environment for artillery is not demanding as the maximum rates are not expected to exceed $100^\circ/\text{s}$. The angular accelerations under firing conditions may, however, be in excess of $300^\circ/\text{sec}^2$.

System Configurations

A proposed form of AGLS is shown in Figure 8. The inertial system or Primary Attitude Reference System (PARS) is mounted on a bracket assembly which, in turn, is secured rigidly to one of the gun trunnions. The attitude data thus relates to a stable reference axis and minimises the uncertainties due to barrel flexure. The PARS provides the servo control system with angular position and rate information when requested by the GLC, which serves as the bus controller. In the RA the 'target' bearing and elevation, as computed by the off-board Command Post, are received by the SPG over the BATES (Battlefield Artillery Target Engagement System) and are transferred to the servo system in readiness for the GLC command to close the position and rate control loops and lay the gun. In the mobile mode, with the elevating mass locked to the hull, the PARS operates as a navigation system, utilising the odometer data in the appropriate way. The data bus configuration is advocated for the AGLS to allow for eventual expansion to accommodate the integration of on-board ballistic sensors and the full computation of 'target' data.

Subject to availability and cost of appropriate solutions, the reversionary mode may be achieved with the use of a Secondary Attitude Reference System (SARS), a lower performance inertial reference, mounted on the same bracket assembly and designed to provide an immediate reversionary capability in the event of a PARS failure. It may be that the SARS would be switched on in parallel with the PARS and would be

continuously updated by it. From the point of PARS failure the SARS might be required to maintain the primary mode of performance for, typically, one hour of gun control without vehicle movement. This may be achievable by implementing a method of differentiation between gyro drift and gun movement and by thus ignoring the former when it is sensed that no gun movement is taking place. The SARS is also intended as a reversionary navigation system in these circumstances. Although certain contenders for the SARS are under consideration it is felt that this application may be more appropriate for a mature fibre optic gyro system, exhibiting a more acceptable cost-performance ratio. A reversionary mode at this time would seem to be the use of the SPG direct fire sight or other optical equipment which could be referenced by the inertial system prior to failure or which could be used with turret and trunnion encoders to obtain azimuth (and elevation) reference from another SPG.

The proposed OPV concept (Figure 9) would once again employ a data bus configuration to allow modularity and future expansion. As discussed previously the inertial system may be mounted on the mantlet (or elevating mass) where it is optimally referenced to the sensors, but where the navigation performance may be compromised by the frequent turret movement. It may be mounted elsewhere in the turret should space constraints dictate, but this would appear to be least attractive since neither attitude nor navigation is optimised. Finally, and perhaps most appropriately in view of the TLE domination by the OPV position accuracy, the inertial system may be mounted in the hull thus optimising navigation performance.

The procedure for the OPV is to observe the target using the appropriate surveillance device and obtain the bearing, elevation and range to it using the inertial system and laser range finder. With the knowledge of the grid position of the OPV, the CDU is able to compute the target grid co-ordinates and transmit these over the BATES to the Command Post.

FIELD ASSESSMENT

It is clear from the foregoing that the SPG application represents a uniquely critical environment for the inertial system. As such, particular care should be taken here in the measurement of INS performance. Some agencies have advocated a technique whereby the gun barrel orientation is measured using theodolites and electronic distance measurement equipment to an accuracy in the order of 0.1 mil. Whilst this method will provide a gun system performance value, care should be taken in the judged apportionment of the inertial contribution to it. Much experimental work has been conducted in the field of barrel flexure due to thermal effects arising from ambient and firing conditions.

Depending on the barrel datum used for such measurement, it may be concluded from this work that such induced errors, together with intermediate mechanical alignment errors, may easily overwhelm those sought from the inertial system itself. The preferred method, currently being refined by the Royal Armament Research and Development Establishment, Fort Halstead, is to mount an optical cube directly onto the INS mounting and to auto-collimate to it thus providing what RARDE claim to be a measurement accuracy of 0.1 mil for the attitude of the INS unit on the gun trunnion.

The selection of navigation routes plays an important role in the assessment of the INS positional accuracy. In particular, closure of a route is to be avoided since this may nullify a number of errors exhibiting heading dependency. For the strapdown INS employing a ZUPT facility, care should be taken in the measurement (carried out statically for practical reasons) of what may be intended to represent intermediate points along a longer route. Unless the ZUPT is disabled under such circumstances, each intermediate point will be refined automatically by the system when the vehicle halts. Depending on the timing and method of measuring both the position and attitude, such a facility could provide a misleading measure of navigation and drift performance.

INTEGRATION WITH GPS

A stand-alone GPS capability is seemingly not viable for these applications which require a high level of availability and pointing accuracy. Two areas of immediate interest, however, are the possible use of a lower grade inertial system in conjunction with GPS (introducing the need for orientation improvement) and the offer of a reversionary capability.

Although GPS/INS mixing is studied throughout the navigation community the subject of orientation improvement is not generally of prime concern. Instead greater attention is normally focussed on reducing position and velocity errors. If used in concert with a high quality strapdown INS employing opportunistic ZUPTs it is doubtful whether appreciable benefits will arise in these areas for the artillery application in particular or indeed for land systems generally.

GPS would provide the artillery with a ready mechanism for automatic initialisation and updating of the INS thus reducing manual data inputs and the dependency on pre-surveyed points which represent, by definition, an operational constraint on the weapon elements concerned.

CONCLUDING REMARKS

The paper has described a unified approach to the analysis of the benefits and performance requirements associated with the on-board integration of INS within two particular inter-dependent elements of the artillery. It has demonstrated that both the SPG and OPV require a comparatively high quality INS and that their respective performance levels are commensurate with a standardized approach.

The rapidly developing maturity of strapdown RLG systems provides these elements of the artillery with a particular potential to realise a major and highly cost-effective operational enhancement in terms of reliability, survivability, reduced ammunition expenditure and manpower/equipment savings.

Emphasis has been placed on the need for careful trials and assessment both to avoid the selection of the wrong technology and/or the misuse of it. Such assessment should perhaps be conducted as part of an overall strategy aimed towards the form, fit and function (F³) standardization of the INS category concerned, thereby permitting competitive procurement along with cost-effective adoption of the standard by further applications as they emerge within the land area.

REFERENCES

1. Ameen, Y.K. and Symonds, G.B., "The Application of On-board Inertial Systems within the U.K Royal Artillery", IEEE PLANS '86 RECORD, November 4-7, 1986, pp. 81-88.

ACKNOWLEDGEMENT

The work reported in this paper is based on the authors' participation in the Artillery Systems Research Programme (ASRP) of Royal Armament Research and Development Establishment, Fort Halstead. In the case of Dr Ameen, this was as a result of MOD contracts awarded to Applied Science Analytics Ltd whilst Mr Symonds is a former manager of the ASRP.

Copyright © Controller HMSO, London 1988

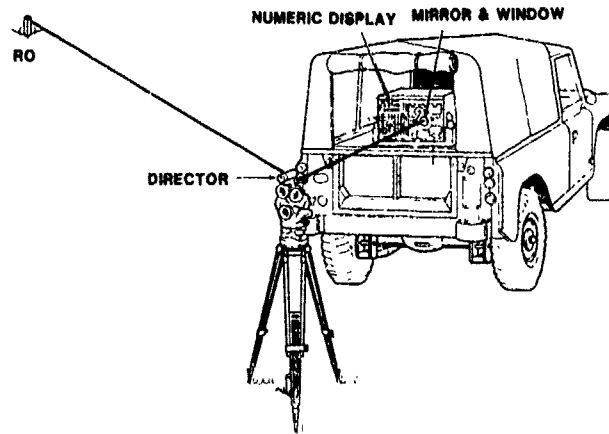


Fig. 1 Battery Survey using PADS

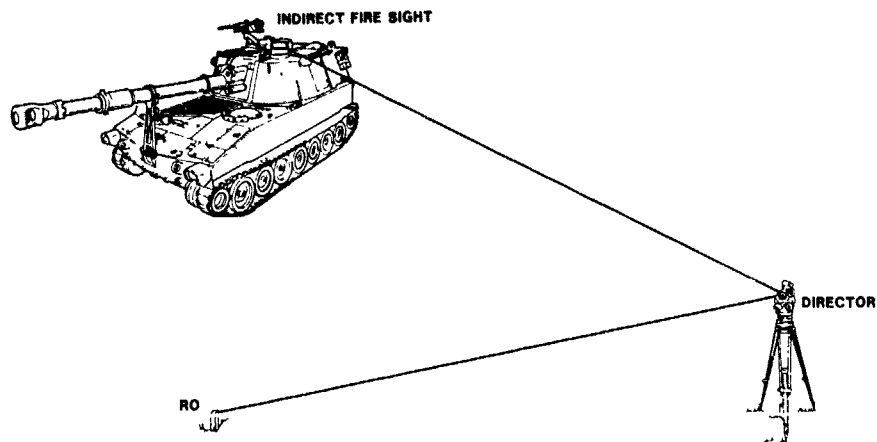


Fig. 2 Passing Line to Gun

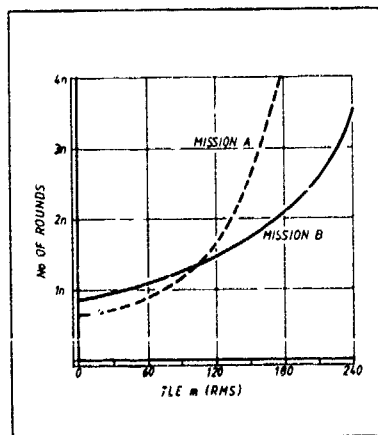
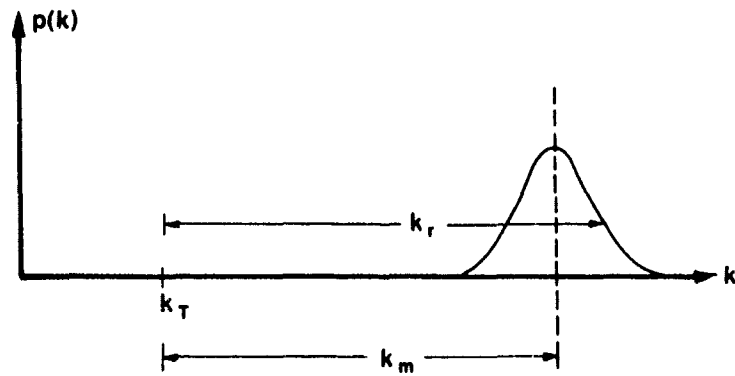


Fig. 3 Relationship between Ammunition Expenditure and Target Location Error



TRUE ODOMETER SCALE FACTOR (OSF) = $S_0 (1 + k_r)$
 WHERE S_0 = NOMINAL OSF VALUE

k_r = ERROR ON THE BASIS OF A SINGLE CALIBRATION RUN

k_m = MEAN ERROR IN THE OSF IF A LARGE NUMBER OF CALIBRATION RUNS ARE PERFORMED

Fig. 4 Typical Distribution of Measured OSF Error for a Given Terrain

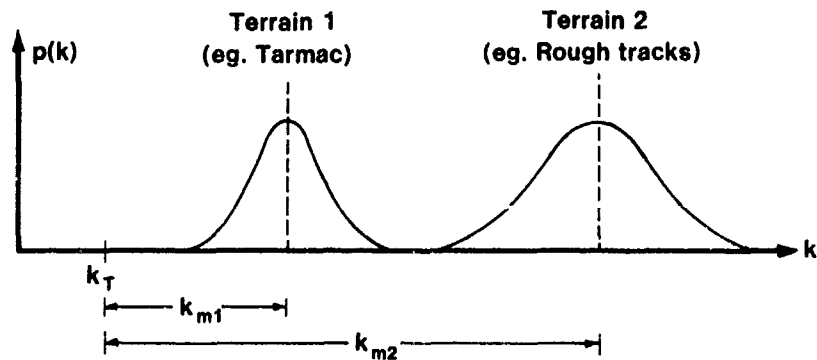


Fig. 5 Typical Distribution of Measured OSF Error for Two Different Terrain Types

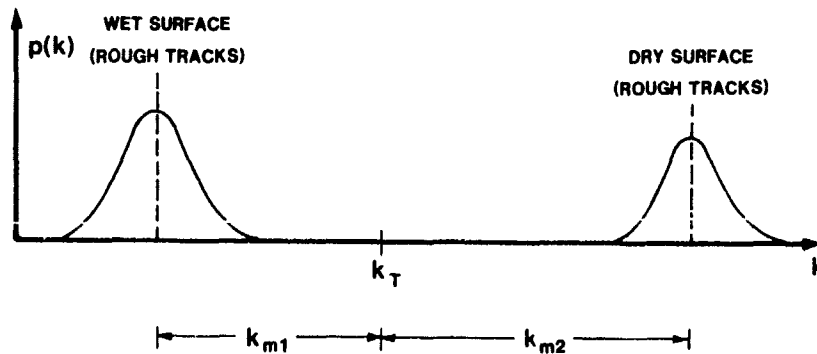


Fig. 6 Typical Distributions of Measured OSF Error for Different Surface Conditions

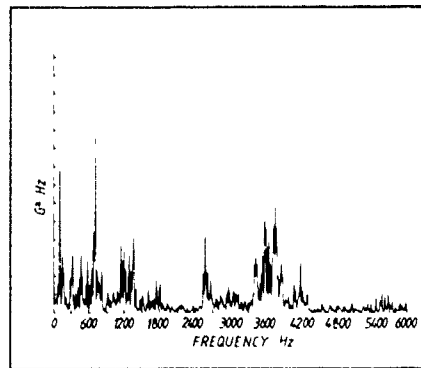


Fig. 7 Shock Spectral Density at the Gun Trunnion (M109 A3 at Charge 8)

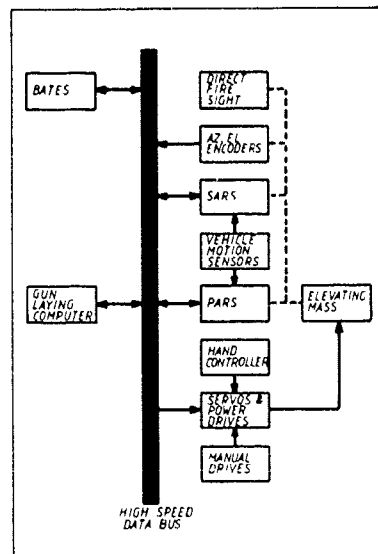


Fig. 8 Schematic Diagram of AGLS

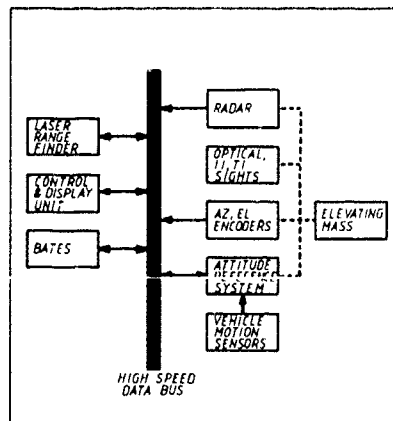


Fig. 9 Schematic Diagram of OPV System

GUIDANCE & CONTROL TECHNIQUES FOR AIRBORNE TRANSFER ALIGNMENT AND SAR MOTION COMPENSATION

by

James L. Farrell
Westinghouse Defense and Electronics Systems Center
PO Box 746
Baltimore, MD 21203
United States

ABSTRACT

A synthetic aperture radar employing an electronically steerable array antenna is simulated for the case of an extant master inertial measuring unit (IMU) in a remote location. To obtain reliable motion compensation data a strapdown IMU is mounted within about a foot of the antenna. Ramifications include (1) the approach to transfer alignment, (2) nav update for the master IMU, and (3) effects of motion-sensitive strapdown inertial instrument degradations on synthetic aperture radar (SAR) mode performance. These ramifications are addressed herein; although results are presented for simulation under simplified conditions only, the path has been prepared for full operational validation as discussed herein.

INTRODUCTION

Most of the growing body of literature involved with synthetic aperture radar (SAR), motion compensation (MC), and transfer alignment deals primarily with conventional mechanizations (i.e., having gyros with spinning rotors), governed by conventional wisdom (i.e., with a "low cost" strapdown unit providing short-term MC information while receiving long-term correction data via transfer alignment from a gimbaled master). RLG implementations will heavily influence the approach, and the ramifications have not yet received sufficient attention. Immediately it needs to be pointed out that the implications cannot be fully described in the open literature. Still there is much to be covered, even in a discussion restricted to functional methodology. While maintaining purely general orientation then, this paper is intended to address significant (but not widely acknowledged) aspects of the operation.

After a brief description of the basic MC function, transfer alignment is addressed from the standpoint of various alternatives, mechanizations, and conventions. A simulation is also described, for which results are then briefly shown under restrictive circumstances. Extension to validation is then addressed, in many forms that include verification of (1) algorithm functional design, (2) operational software code, emulated on a mainframe host, (3) digital hardware on the bench, and (4) combinations of these.

MOTION COMPENSATION (MC)

Processing of incoming SAR data imposes the task of correcting rf phase. To support this function the unit line-of-sight vector U is repeatedly updated based on the strapdown velocity V expressed in locally level (e.g., wander azimuth) reference coordinates. The time integral of V is augmented by the lever arm displacement vector d from strapdown inertial measuring unit (IMU) to antenna phase center, transformed through the 3×3 orthogonal matrix $[T]$ of direction cosines from airframe to nav coordinates thus: Instantaneous position vectors of the radar R and strapdown IMU R_s are related by

$$R = R_s + [T] d$$

so that their velocity vectors are related essentially by

$$\dot{R} = \dot{R}_s + [T] \{W\} d$$

where $[W]$ is a skew-symmetric matrix of angular rates. The time integral of the along-range component of \dot{R} is obtained by recursively accumulating the sum, over each interpulse period,

$$S(N) = S(N-1) + (U^T \dot{R}) t_{pp} + U^T \int \{ [T] \{W\} d \} dt$$

which, in view of the direction cosine matrix updating relation, reduces to

$$S(N) = S(N-1) + (U^T \dot{R}_s) t_{pp} + U^T [T] d$$

where $[T]$ is now evaluated at a specified instant within the interpulse period (t_{pp}). Range gates are shifted in accordance with the ratio of $S(N)$ to range resolution, and the phase correction $\approx [4 \pi S(N) / \text{wavelength}]$. LOS direction also enters the beam steering controller for antenna pointing.

Because the nav data processing is not slaved to the radar, the time that elapses between data measurement and its use by SAR mode cannot be known *a priori*. Thus time tags are assigned to INS data, and delays between measurement and usage of data call for extrapolation to the instant when used. Straightforward data fitting has been used successfully for this purpose.

TRANSFER ALIGNMENT

Precise correction of rf phase dynamics imposes stringent demands on the MC subsystem. The short term dynamics experienced by the antenna must be monitored by a strapdown INS (an alternative, consisting of an accelerometer triad without strapdown gyros, is dismissed in APPENDIX A). It is well known that the strapdown inertial measuring unit (IMU), while providing excellent response to the directional variations of its airframe mounting site, tends to report its motion increments with an accompanying gradual drift in its apparent reference frame orientation; thus a gimbaled master INS is used for transfer alignment. The transfer alignment algorithm restores coincidence between (slowly varying) gimbaled and strapdown reference frames while allowing the strapdown IMU to recognize legitimate differences in motion and attitude at different airframe stations.

The transfer alignment formulation chosen has much in common with Refs. 1 and 2 (although the former describes alignment between two gimbaled platforms). Since the gimbaled master is the trusted long term attitude reference, its inner element will be the desired strapdown reference at all times. There are additional ramifications to this approach:

1) For transfer alignment operation, the master is considered perfect while strapdown alignment error is driven by relative drift (i.e., difference between strapdown and master INS drift rates). For SAR performance analysis purposes, on the other hand, master and strapdown drift effects are additive. A related principle applies to vertical deflections; they have no effect on transfer alignment, but drive overall nav error in the usual way (i.e., as a forcing function, superimposed on accelerometer bias effects, in the dynamic equation for velocity error).

2) The master INS itself can receive nav update information, which will affect the slave gradually, via transfer alignment. It is recognized that, in some respects, (e.g., to reduce settling time), direct strapdown alignment via navaid in tandem with master update would be preferable. The time alignment implementation, however, would compromise that benefit, from the standpoint of performance as well as complexity.

3) Here, as elsewhere, strapdown transfer alignment error is defined as the offset between the computed strapdown reference frame and the actual orientation of the master platform inner element. This latter item adheres to the popular concept of "aligning" two reference frames while of course maintaining the necessary distinctions from other angular deviations such as (a) mounting misalignments (pp. 72 and 111 of Ref. 3), (b) vertical deflections (which, as explained in Ref. 2, do not degrade transfer alignment), (c) shock mount deformation, which is equivalent to an additive error, and (d) aeroelastic vibration (relative motion between airframe stations used as master and strapdown mounting sites, which is of course a dominant justification for employing a strapdown package).

The foregoing considerations facilitate both the identification of central issues and the selection of an approach that meets all needs. A primary issue is the frequently encountered problem of full observability. Its first implication involves the familiar deficiency of azimuth information; whereas a sustained lift force readily provides adequate information for leveling, prompt azimuth correction from INS-derived translational data alone requires horizontal components of acceleration (Ref. 4). In some applications the mission profile contains sustained turn segments which provide those horizontal acceleration components; in this operation, however, that plants the seeds for a host of additional potential motion-sensitive degradations. Chief among these, in conventional strapdown mechanizations, are scale factor and mounting misalignment errors for each gyro. What makes these particularly troublesome is their capacity for destroying the short-term stability of the velocity accuracy; that loss is the Achilles' heel of conventional strapdown mechanizations. Attempts to determine these effects via augmenting states, however, further compounds the observability problem. The APPENDICES address these issues in greater detail; the outcome of weighing all factors is a selected approach that minimizes the number of states while employing no angular rate matching.

The next decision made in defining the transfer alignment Kalman filter was the selection of state variables. Actually a major step in that direction had already been taken, i.e., to require commonality between transfer alignment and in-flight nav; see APPENDIX B. Thus if the 3×1 vectors $[\psi, v, A, p, n, N, e]$ respectively denote strapdown misalignment and velocity error, specific force, flight profile rate, gyro drift state, and random excitation for translation and drift states, all expressed in the reference (master platform inner element) coordinate frame, then, from Eq. (3-37) of Ref. 3,

$$\dot{v} = \psi \times A + n$$

while, for Kalman filter data averaging intervals $\ll 84$ min., the first term on the right of Eq. (6-20) of Ref. 3 can be dropped so that the misalignment dynamics simplify to

$$\dot{\psi} = -p \times \psi + N$$

under conditions of short-term quasistatic drift ($\dot{N} = 0$) so that, with state and random excitation vectors denoted as

$$x = \begin{bmatrix} v \\ \psi \\ N \end{bmatrix}; \quad w = \begin{bmatrix} n \\ 0 \\ e \end{bmatrix}$$

all of the above relations can be compressed into the standard form

$$\dot{x} = [A]x + w$$

where, with 3×3 null and identity matrices represented by $[0]$ and $[I]$,

$$[A] = \begin{bmatrix} [0] & (-A \times) & [0] \\ [0] & (-p \times) & [I] \\ [0] & [0] & [0] \end{bmatrix}$$

Several important items remain to be discussed; for brevity they will be covered by a simple enumeration:

1) The data averaging time of the Kalman filter will, via spectral density levels for the excitation in Eq. (5-57) of Ref. 3, be set as a function of the Kalman filter data averaging intervals, e.g., only seconds for leveling (Ref. 4).

2) Within these short intervals, accelerometer bias effects cannot be distinguished from initial misorientation effects without an impracticable maneuver sequence. Furthermore, accelerometer bias observability problems would have been compounded by myriad unmodeled errors when a strapdown slave IMU is used. This reasoning is consistent with avoidance of accelerometer bias states.

3) The above reasoning applies even more strongly to avoidance of augmenting states for gyro scale factor error and mounting misalignments.

4) Formation of master-slave velocity residuals requires an "omega cross-lever-arm" velocity term. The requisite angular rate information is available from the strapdown IMU. Angular rate noise degradation could be reduced, as in Ref. 1, by adding three position states and substituting position changes for velocity difference measurements. To keep the state dimensionality at a minimum, however, this option is not adopted here.

5) As with any estimation algorithm, residual dynamics can be monitored and filter gains reinitialized by resetting the covariances at any time.

6) Strapdown updating and alignment algorithms can be implemented with a "hands off" approach, wherein data flow is unidirectional from the strapdown computer outward, and prescribed adjustments to computed attitude and velocity are maintained in separate computations. This is the approach adopted here, e.g., estimated corrections are not allowed to adjust the strapdown direction cosine matrix. The reason involves the previously mentioned task of time tagging and artificial synchronization in the presence of multiple asynchronous information sources.

The synchronizing operation just mentioned exerted considerable influence on the entire approach. The mundane task of data time-alignment is never as simple in operation as it is conceptually, and it is not quite standardized. Still it has been solved many times and, with each added success, the implementation becomes more manageable.

SIMULATION

Design cannot rely exclusively on intuitive analysis nor, at the other extreme, on prohibitively expensive flight testing. As an intermediate step, prospective algorithms should be computer-tested with input data sequences in rigorous conformance to time histories that would result from controlled repeatable conditions. An extremely flexible structured program has been developed for that purpose, wherein performance for a variety of algorithms and parameters can be evaluated for any flight path, with various arrays of radar scatterers, in the presence of virtually any reasonable combination of degradations. When all these error sources are nulled, acceptability of a processing sequence is easily evaluated by judging results versus known correct outputs. For example, a nominal 40-ft resolution case with sidelook geometry for straight level flight at 400,000 ft range has well known expected response characteristics. With 0.08-microsecond pulses spaced a millisecond apart, a 2048-point Hamming-weighted FFT performed at the outputs of active motion-compensated range gates produces separable responses with computed coordinates ("X" and "Y") correct to within a half-resolution cell, and with acceptable spreading in proximate cells (five cells on either side of the cell having maximum response in each range gate) as shown. Also, the normalized sum of squares of responses from all other (2048-1-10=2037) remote cells ("GUDGE") represents an acceptable integrated side lobe ratio (ISLR).

TSTEP = 0.0010000000 TSTOP = 2.0481

GATE	NFMAX	X	Y	dB	GUDGE
1	0	*****	****	*****	*****
2	1841	-9990.	-38.	16.8	44.6
3	1941	-9992.	11.	0.0	38.0
4	0	*****	****	*****	*****
5	409	10018.	129.	3.1	36.6
6	409	10019.	178.	7.8	39.4
7	0	*****	****	*****	*****

PROXIMATE CELL RESPONSES (dB ATTENUATION)									
5	4	3	2	1	1	2	3	4	5
****	****	****	****	****	****	****	****	****	****
39	38	41	31	7	3	21	43	39	39
49	49	50	49	9	4	31	61	49	49
****	****	****	****	****	****	****	****	****	****
44	45	58	25	3	10	14	46	44	45
42	43	51	22	3	10	43	44	42	43
****	****	****	****	****	****	****	****	****	****

FFT CELL NUMBERS:										
5	4	3	2	1	0	1	2	3	4	5
2043	2044	2045	2046	2047	2048	1	2	3	4	5
1636	1637	1638	1639	1640	1641	1642	1643	1644	1645	1646
1636	1637	1638	1639	1640	1641	1642	1643	1644	1645	1646
2043	2044	2045	2046	2047	2048	1	2	3	4	5
404	405	406	407	408	409	410	411	412	413	414
404	405	406	407	408	409	410	411	412	413	414
2043	2044	2045	2046	2047	2048	1	2	3	4	5

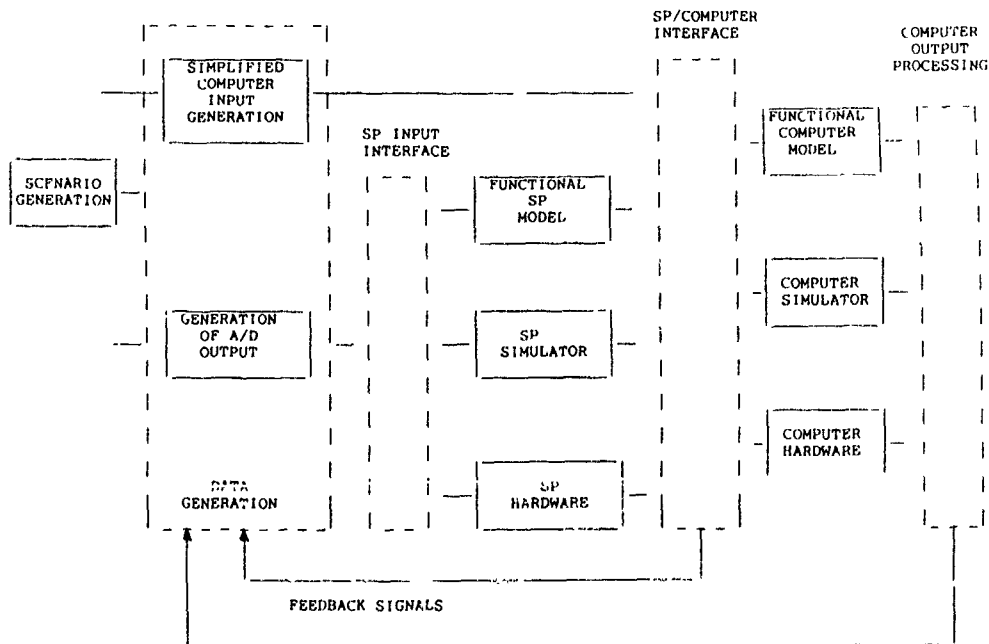
Further validation of the algorithms can take an almost unlimited variety of forms, such as insertion of degradations described in Ref. 5, separately or in combination. Actually these sources are present in the simulation (and consistent with the descriptions in Chapter 4 of Ref. 3), but obviously only a brief example can be given here. To illustrate, the next run inserts an acceleration bias of one wavelength per second squared (which in this case produces one cycle of quadratic phase error; it could have come from uncorrected transfer alignment error, accelerometer bias, vertical deflection, drift effects accumulated within the frame, etc.). Proximate cell responses were significantly affected, with spreading in the dominant response gates (shown with arrows below) conforming to theoretical levels.

Simulation represents only the first of four operating modes, which will allow any combination of simulated or actual input data and simulated or actual airborne processing. Thus the complete set of four modes includes

- (1) simulation (simulated input data and simulated processor algorithms),
- (2) validation (simulated input data and actual processing code),
- (3) post-flight (actual input data and simulated processor algorithms),
- (4) flight test (actual input data and actual processing code).

For SAR evaluation, the model used herein allows 1) dynamic motion characteristics conforming to any realistic flight conditions, 2) full flexibility in placement of ground reflectors, 3) adherence to timing and quantization effects to be introduced through digital processing, 4) motion compensation errors arising from INS imperfections, and 5) complete generality in geometry assumed for repeatable test scenarios. Fortunately these features constitute essentially an air-to-ground counterpart for air-to-air programs already designed and verified. The input data driver is based directly on Ref. 8, while digital assembly test provisions conform to those described in Refs. 9 and 10; combination of these in modular fashion is consistent with the approach described in Ref. 10.

The arrangement shown in the accompanying figure (taken from Ref. 10) demonstrates the high degree of flexibility provided for radar mode software validation. The radar computer (RC) can be represented by (1) a functional model of its algorithms, (2) a step-by-step emulation of its operations, driven by operational flight program (OFP) code while executed on a general purpose (GP) host, or (3) RC hardware, if available. Computer inputs can come from (1) direct generation of simulated signal processor (SP) outputs, (2) a functional model of the SP, (3) an SP simulator analogous to the previously mentioned step-by-step emulation of computer operations, or (4) SP hardware, if available. This scheme is clearly amenable to both SP/RC compatibility testing and separate validation of SP or RC software, while also relatively invulnerable to delays caused by unavailability of digital hardware (and of course, totally independent of all hardware other than the digital assembly).



DESIGN AND VALIDATION OPTIONS

APPENDIX A: NEED FOR A FULL STRAPDOWN INSTRUMENT PACKAGE

It remains to be explained why an accelerometer triad without gyros cannot perform the MC function. Briefly, the specific force vector F experienced by the radar is linked to that of the strapdown IMU F' via a relation in radar-station airframe coordinates expressible to first order as

$$F = F' + f + e \times F'$$

where f denotes relative (radar/IMU) acceleration and e essentially represents difference in attitude at the two stations on the airframe. Were it not for this last term, antenna-mounted accelerometers would contain just the information needed to correct for the relative motion. Customarily an attempt is made to separate the "low" frequency (attitude difference) from the "high" frequency (f) effects; obviously an appropriate "crossover" can be found only if these effects are in fact spectrally separable. In general they are not, so that spectral overlap between $e \times F'$ and f plagues that approach.

APPENDIX B: CHOICE OF FORMULATION

If Ref. 1 had not already established a successful precedent (albeit for two gimbaled platforms), the formulation chosen here might have been considered a significant departure from conventional wisdom; Refs. 1 and 2 chose angular *uncertainties* about *reference* axes, in marked contrast to the *actual* angular displacements between *vehicle* axes used in Ref. 6 and 7. An obvious reason for that decision was avoidance of computational duplication; the formulation in Ref. 2 retains the full benefit of all strapdown data processing (including on-line instrument compensation as well as high-speed attitude updating) already done elsewhere in the avionics, and the master/strapdown misalignment states are slowly varying. Subsequent work performed by the authors of Ref. 7 now show a similar preference for this choice of misalignment states, but another characteristic of many CSDL reports on coordinate alignment (disposition of velocity states) remains: It is entirely permissible to omit velocity vector components from the state formulation and express master-slave velocity residuals in terms of integral functions, *but* to do so is to forfeit commonality with INS updates as usually formulated. Thus Refs. 1 and 2 exhibit another major departure from the CSDL approach, by characterizing the transfer alignment dynamics in the same manner as short-term INS error propagation — again without incurring any disadvantages.

APPENDIX C: DISPOSITION OF ROTATIONAL OBSERVATIONS

Relative motion between airframe stations used as master and strapdown mounting sites is of course the whole justification for employing a strapdown package. Nevertheless it is not uncommon to hear mention of "misalignments" between *airframe* orientation extant at the two mounting sites. Although that assessment of rotational vibrations is indisputably correct in a literal sense, the vocabulary must not be allowed to obscure the identity of the alignment error (between *reference* frames) to be corrected by the updating methods described here.

The above clarifications are often necessary to ensure proper interpretation of what has evolved as accepted terminology in this field. For example, the expression "attitude matching" if taken literally could imply Kalman filters designed to drive the strapdown attitude solution toward the instantaneous attitude indicated by the gimbaled master. Clearly that is not the intent of the operation; it would be tantamount to suppressing the very information that the strapdown INS is intended to monitor. Furthermore, even "angular rate matching" is not literally correct in this application; the angular rate vector is uniform throughout all points on a rigid structure *only*. Admittedly, information can be extracted from finite rotations *but*, to apply that procedure to the nonrigid airframe, certain additional burdens would have to be accepted; these are briefly discussed below.

Potential benefits of rotational observations can first be visualized in terms of an ideal snap roll, performed by a rigid aircraft and monitored accurately by pre- and post-maneuver gimbal readings from its master platform. The same angular excursion would then be observed by the strapdown INS, in general resolved along a noncoincident triad of axes. If the strapdown reference frame had drifted off the gimbaled master inner element orientation (by an amount denoted here as ψ radian, with acceptable values typically below a milliradian), the effect of that deviation would be present in a signal proportional to ψ (e.g., with magnitude of up to 1.5 mr for ~ 90-deg snap roll). The signal would be accompanied by a degradation, however, which would include at least the platform gimbal pickoff uncertainty (also on the order of a milliradian). Furthermore, unless augmenting states were added to the transfer alignment formulation, the degradation would be aggravated by strapdown roll gyro scale factor error combined with gyro mounting misalignments (i.e., unknown deviations of each gyro input axis from the corresponding accelerometer-based instrument package coordinate system). These errors would limit the gain that could legitimately be used to process the rotational observation in the transfer alignment Kalman filter (and thus limit the effectiveness of these observations). Compounding the problem, rotational motions excite a host of additional dynamics-sensitive strapdown instrument errors; but any attempt to reduce the maneuver excursion would reduce the signal and thereby further compromise the effectiveness of this rotational information. Now drop the assumption of rigidity, and visualize the signal accompanied also by an *actual* difference in rotational excursion experienced. The approach to this complication in Ref. 6 is to treat structural deformation effects as "unwanted information or noise" in the characterization of rotational observations (which implies a literal interpretation of the term "angular rate matching"). That method will reduce the signal-to-noise ratio, however, to levels below $[\psi / (\text{angular deformation})]$. In applications wherein the aeroelastic/structural deformation can substantially outweigh allowable coordinate alignment error, the result would be an unacceptably long settling time for the transfer alignment Kalman filter. An alternative procedure for nonrigid structures is to include deformation states in the dynamic formulation, as was done in Ref. 2. That step was taken as a last resort, however, for an application wherein azimuth observability could not be guaranteed by any other means.

Only after all of these ramifications are clearly faced can the prospects for augmenting the state by gyro mounting misalignments and scale factor errors begin to be seen in proper perspective. Moreover, that approach immediately adds nine more states to be determined — and effects of those nine are not readily distinguishable from *each other*, let alone from the effects of the three unaugmented misorientation state variables. As if that were not enough, multiple accompanying unmodeled motion-sensitive degradations force whatever states are modeled to be

incorrectly estimated; those states receive the "blame" for the unmodeled effects. Inevitably this means that, whenever the dynamics change, the transfer alignment has to be redone to accommodate the new dynamic conditions. But the plurality of states in the augmented formulation ensures long settling times as well as observability problems.

The weight of all these factors explains much about this author's (1) concerns regarding rotational measurements and observability in general, (2) preference for formulations with minimal state dimensionality, and (3) enthusiasm for RLK implementations, which virtually eliminate gyro scale factor errors and several additional motion-sensitive degradations.

REFERENCES

1. Yamamoto and Brown, "Design, Simulation, and Evaluation of the Kalman Filter Used to Align the SRAM Missile," AIAA Paper No. 71-948.
2. Farrell, "Transfer Alignment for Precision Pointing Applications," NAECON, Dayton Ohio, 1979.
3. Farrell, *INTEGRATED AIRCRAFT NAVIGATION*, Academic Press 1976.
4. Bryson, "Rapid In-Flight Estimation of IMU Platform Misalignments Using External Position and Velocity Data," Air Force Contract Final Report #AFAL-TR-73-288, Sept. 1973, pp. 19-24.
5. Farrell, "Strapdown INS Requirements Imposed by SAR," AIAA JG&C, Jul-Aug 1985.
6. Schultz and Keyes, "Airborne IRP Alignment Using Acceleration and Angular Rate Matching," *JACC Proceedings*, Ohio State Univ., Columbus, June 1973.
7. Harris and Wakefield, "Coordinate Alignment for Elastic Bodies," NAECON, Dayton Ohio, 1977.
8. Hedland and Farrell, "Simulation of Tracking Radar in the Presence of Scintillation," NAECON 1980.
9. Horner, "An Embedded Software Development and Integration Facility," NAECON 1981.
10. McSweeney, Farrell, and Hedland, "Integrated and Transferable Hardware/Software Checkout," NAECON 1982.

STATE ESTIMATION FOR SYSTEMS MOVING THROUGH RANDOM FIELDS*

by

Donald E. Catlin
Associate Professor of Mathematics
University of Massachusetts
Amherst, Massachusetts 01003
United States

and

Robert L. Geddes
Member of Technical Staff
The Analytic Sciences Corporation
55 Walkers Brook Drive
Reading, Massachusetts 01867
United States

SUMMARY

A growing memory discrete dynamic model for performing temporal extrapolations along a predetermined path in a random field is presented in this paper. This dynamic model is used to drive a linear system that is itself driven by discrete white noise. The coupled system is used to derive a state estimation scheme that recursively processes noisy measurements of the system. In addition, using the aforementioned dynamic model as a reference (truth) model, the authors develop a covariance analysis to measure the estimation errors that occur when the dynamics along the path through the field are modeled as a Markov linear model and state estimation is performed using discrete Kalman filtering. The performance evaluation of an inertial navigation system influenced by the earth's gravity field aboard a maneuvering ship is provided as a specific illustrative example.

1 INTRODUCTION

Inertial navigation system (INS) accuracy, most notably velocity accuracy, depends upon the accuracy of the compensation used by the INS for specific force due to gravity [3]. Differences between the local gravity vector and the implemented gravity compensation give rise to a vector orientation error (vertical deflection) and a magnitude difference (gravity anomaly). These "gravity errors" are, in general, spatially correlated and are typically modeled as Markov processes when Kalman filtering is used to estimate and compensate INS errors [5]. When the vehicle motion is along a great circle, this Markov assumption is reasonable since one would expect gravity errors to decorrelate in time. However, such a model is conceptually wrong in the presence of turns since the correlation distances will be "perceived" by the filter as being greater than they really are. Further, across-track vertical deflections can become along-track deflections and conversely, the vehicle could return to a previously occupied position, and so on. The situation is further exacerbated when gravity measurements (using a gravity gradiometer and/or a gravimeter) are processed because the measurement model used in a suboptimal Kalman filter is not matched to the actual spatial model. It is therefore desirable to develop a state estimation scheme using a gravity model that is consistent with the underlying spatial gravity field.

The above problem has been addressed by other authors ([1], [4], [7], [9]) who have either obtained approximate solutions or have obtained complete solutions for selected restricted maneuvers. The most recent paper [9] also presents an analysis of system covariances under the assumption that the stochastic field is spatially separable.

In this paper, the above problem is formulated as a discrete linear system being driven, in part, by movement through a random field. In the case of a navigation platform, this has the effect of assuming the vehicle motion is a sequence of straight line movements; i.e., the track is a "polygonal line." It is further assumed that the along-track extrapolation through the field is a linear combination of field states and noise. In Section 2, it is shown that extrapolation can be carried out in a manner which is consistent with these

*This paper appeared in the IEEE Transactions on Aerospace and Electronic Systems, as "State Estimation and Divergence Analysis," Vol. AES-20, No. 5, September 1984, ©IEEE 1984.

assumptions and the method of doing so is unique. To obtain this extrapolation it is necessary to assume that the covariance between any two points in the field is known. However, no other description of the field is required. This extrapolation model is then coupled with the model for a discrete dynamical system, (e.g., a navigation error model) and in Section 3 this combination is used to develop an optimal recursive state estimator which is very similar to a Kalman filter. In Section 5 this same combined model is used as a reference (truth) model to develop the equations for a divergence analysis of a Kalman filter operating in the non-Markovian setting described by the reference model.

An essential feature of the above scheme is that it contains a growing memory. This gives rise to two distinct but related problems. First, it is necessary to continually calculate the pseudoinverse of a growing covariance matrix. In Section 4, this problem is addressed and a scheme is presented whereby this pseudoinverse can be calculated recursively, thereby reducing the computational burden. The second problem is memory storage space. Although this problem is not addressed herein, the results in Section 4 can be interpreted to obtain a systematic algorithm for selectively deleting past information, thereby producing a suboptimal filter with bounded storage requirement. The question of which information should be deleted is being studied.

It should be pointed out that the analysis technique below applies to any linear system driven by both white noise and motion through any random field; the navigation application is presented for purposes of motivation.

2 LINEAR EXTRAPOLATOR

Let (Ω, P) be a probability space and X an arbitrary set. By a random vector field on X is meant a function

$$f: X \times \Omega \rightarrow R^n$$

Thus, for each $x \in X$, $f(x, \cdot)$ is a random vector. Although it is customary to equip X with some structure, e.g., make X a differentiable manifold, in this treatment X can remain structureless. For convenience define a function u as

$$u(x) \triangleq f(x, \cdot)$$

thereby suppressing the second element. It is assumed that if x and y are any two points in X , the correlation function

$$\phi_{xy} \triangleq E[u(x)u(y)^T] \quad (2-1)$$

exists and is known

2.1 DEFINITION

A discrete track T of length N_T in X is a finite subset of X , say

$$T = \{x_1, x_2, \dots, x_{N_T}\} \quad (2-2)$$

2.2 DEFINITION

An along-track linear field extrapolator (ATLFE) is a pair (A_T, ξ_T) , defined for each track T , where given k such that $1 \leq k \leq N_T$, $A_T(k)$ is an $n \times n$ matrix and $\xi_T(k)$ is a random n -vector subject to the following conditions:

- $\xi_T(k)$ is uncorrelated in the sense that

$$E[\xi_T(i) \xi_T(j)^T] = 0 \text{ for } i \neq j \quad (2-3)$$

$$\bullet \quad u(x_{k+1}) = A_T(k) \begin{bmatrix} u(x_1) \\ \vdots \\ u(x_k) \end{bmatrix} + \xi_T(k) \quad (2-4)$$

$$\bullet \quad E[u(x_i)\xi_T(k)^T] = 0 \text{ for } 1 \leq i \leq k \quad (2-5)$$

The above restrictions appear to be rather severe in that condition (2-4) is an equality rather than an approximation, and is independent of track length and initial conditions. This implies that if two tracks T_1 and T_2 have the same last point, i.e., $x_{N_{T_1}} = x_{N_{T_2}}$, then both schemes (A_{T_1}, ξ_{T_1}) and (A_{T_2}, ξ_{T_2}) produce the

same extrapolated value at this point. Condition (2-4) also implies that the scheme is consistent with the statistical properties of the field f along any track. Thus, the question of the existence of an ATLFE is an issue that initially must be settled.

In the following discussion, it is assumed that a particular track T is arbitrarily chosen and fixed. It is convenient to use notation that suppresses the track dependence and emphasizes the discrete time dependence. In particular, relations (2-1) through (2-5) are rewritten as

$$\phi_{1j} = E[u(1)u(j)^T] \quad (2-6)$$

$$E[\xi(1)\xi(j)^T] = 0 \text{ for } 1 \neq j \quad (2-7)$$

$$u(k+1) = A(k) \begin{bmatrix} u(1) \\ \vdots \\ u(k) \end{bmatrix} + \xi(k) \quad (2-8)$$

$$E[u(1)\xi(k)^T] = 0 \text{ for } 1 \leq i \leq k \quad (2-9)$$

The following quantities are defined for notational convenience:

$$u(k) \triangleq \begin{bmatrix} u(1) \\ \vdots \\ u(k) \end{bmatrix} \quad (2-10)$$

$$\Phi(k) \triangleq E[u(k)u(k)^T] \quad (2-11)$$

$$\psi(k+1) \triangleq [\phi_{k+1,1}, \phi_{k+1,2}, \dots, \phi_{k+1,k}] \quad (2-12)$$

$$Q(k) \triangleq E[\xi(k)\xi(k)^T] \quad (2-13)$$

The analysis below depends heavily on the Gauss-Markov estimation theorem [6] and its generalization to non-singular covariance matrices, which is stated without proof [8]. This theorem utilizes several matrix properties. If B is any matrix, B'' denotes the orthogonal projection onto the range of B [2], and the

pseudoinverse of B is denoted by B^+ . The properties of B^+ can be found in [6] and [8].

Theorem 2-1 Let $\{y_1, \dots, y_n\}$ be random k -dimensional vectors whose components are square integrable over some probability space (Ω, P) , i.e., each $y_i \in L_2(\Omega, P)^k$. If Y represents the linear span of the components of y_1, \dots, y_n , and if $z \in L_2(\Omega, P)^k$, then the orthogonal projection, \hat{z} , of z onto the subspace Y^k is given by

$$\hat{z} = E[zy^T] E[yy^T]^{-1} y \quad (2-14)$$

where

$$y^T \triangleq (y_1^T, y_2^T, \dots, y_n^T) \quad (2-15)$$

It follows that $\hat{z} - z$ is orthogonal to each y_i .

Theorem 2-2 For the random field f there exists an ATLFE

Proof. For any track T , define a random vector, ξ , and the matrix A via

$$\xi(k) \triangleq u(k+1) - \psi(k+1) \Phi(k)^+ u(k) \quad (2-16)$$

$$A(k) \triangleq \psi(k+1) \Phi(k)^+$$

Condition (2-3) is thus established by definition. It remains to show conditions (2-2) and (2-4).

By Theorem 2-1 and definitions (2-10) through (2-13), $\psi(k+1)\Phi(k)^+ u(k)$ is the orthogonal projection of $u(k+1)$ onto the Cartesian product of the span of the components of $u(1), \dots, u(k)$; and hence the random vector $\xi(k)$ is orthogonal to each $u(i)$, $i \leq k$. This is condition (2-5).

To show condition (2-3), it can be assumed without loss of generality that $i < j$. From Eq. (2-16) it follows that $\xi(i)$ is a linear combination of the vectors $u(1), \dots, u(i+1)$ and so by Theorem 2-1, $\xi(j)$ is orthogonal to $\xi(i)$ since it is orthogonal to each of $u(1), \dots, u(i+1)$.

Theorem 2-3 Let (A, ξ) be an ATLFE. Then for any track T , (A_T, ξ_T) must necessarily have the form shown in Eq. (2-16).

Proof. From Eq. (2-8)

$$u(k+1) = A(k)u(k) + \xi(k) \quad (2-17)$$

Then from definition (2-12) it follows that

$$\begin{aligned} \psi(k+1) &= E[u(k+1)u(k)^T] \\ &= A(k)E[u(k)u(k)^T] + E[\xi(k)u(k)^T] \\ &= A(k)\Phi(k) \end{aligned} \quad (2-18)$$

the last equality following from Eqs. (2-8) and (2-9). Thus

$$A(k)\Phi(k)^+ = \psi(k+1)\Phi(k)^+ \quad (2-19)$$

Now

$$\begin{aligned}\Phi(k)^n \underline{u}(k) &= \Phi(k) \Phi(k)^+ \underline{u}(k) \\ &= E\{\underline{u}(k) \underline{u}(k)^T\} E\{\underline{u}(k) \underline{u}(k)^T\}^+ \underline{u}(k)\end{aligned}$$

and by Theorem 2-1, $\underline{u}(k)$ is the orthogonal projection of itself onto the k -fold Cartesian product of the span of the components of $\underline{u}(k)$. In other words,

$$\Phi(k)^n \underline{u}(k) = \underline{u}(k) \quad (2-20)$$

It then follows from Eq. (2-19) that

$$\Lambda(k) \Phi(k)^n \underline{u}(k) = \psi(k+1) \Phi(k)^+ \underline{u}(k)$$

and by Eq. (2-20)

$$\Lambda(k) \underline{u}(k) = \psi(k+1) \Phi(k)^+ \underline{u}(k) \quad (2-21)$$

This last expression implies that Eq. (2-17) must necessarily have the form

$$\underline{u}(k+1) = \psi(k+1) \Phi(k)^+ \underline{u}(k) + \xi(k) \quad (2-22)$$

Theorems 2-2 and 2-3 imply that there is one and only one ATLFE, namely that specified by Eq. (2-22). Moreover, Eq. (2-22) is then consistent with the first- and second-order statistics of the underlying field f .

3. OPTIMAL LINEAR FILTERING

A discrete linear system driven by both white noise and motion through a random field can be represented by the equation

$$\underline{x}(k+1) = \Theta_{11}(k) \underline{x}(k) + \xi_s(k) + \Theta_{12}(k) \underline{u}(k) \quad (3-1)$$

where \underline{x} represents that state vector, $\Theta_{11}(k)$ is the state transition matrix, $\Theta_{12}(k)$ is a distribution matrix for \underline{u} , $\xi_s(k)$ represents discrete uncorrelated process noise, and $\underline{u}(k)$ is as defined in Section 2. The problem addressed in this section is that of estimating the state vector \underline{x} by making noisy measurements of the form

$$\underline{z}(k) = H(k) \underline{x}(k) + w(k) \quad (3-2)$$

where

$$\underline{x}(k) \triangleq \begin{bmatrix} \underline{x}(k) \\ \underline{u}(k) \end{bmatrix} \quad (3-3)$$

and $w(k)$ represents the discrete uncorrelated measurement noise. In the following analysis it is convenient to introduce a new vector

$$\underline{x}(k) \triangleq \begin{bmatrix} \underline{x}(k) \\ \underline{u}(k) \end{bmatrix} \quad (3-4)$$

whose dimension increases at each discrete propagation step; $\underline{u}(k)$ is defined by Eq. (2-10). Invoking Eq. (2-22), Eqs. (3-1) and (3-2) can be rewritten as

$$\underline{x}(k+1) = \Theta(k) \underline{x}(k) + S(k)^T \xi(k) \quad (3-5)$$

$$\underline{z}(k) = H(k) \underline{x}(k) + w(k) \quad (3-6)$$

where

$$\xi(k) \triangleq \begin{bmatrix} \xi_s(k) \\ \xi_u(k) \end{bmatrix}, \quad (3-7)$$

$$S(k) \triangleq \begin{bmatrix} I & 0 \\ 0 & M(k+1) \end{bmatrix} \quad (3-8)$$

$$M(k) \triangleq [0, 0, \dots, 0, I_k], \quad (3-9)$$

$$H(k) \triangleq H(k)S(k), \quad (3-10)$$

$$\Theta(k) \triangleq \begin{bmatrix} \Theta_{11}(k) & \Theta_{12}(k)M(k) \\ 0 & I(k) \\ 0 & \psi(k+1)\Phi(k)^T \end{bmatrix}, \quad (3-11)$$

and ξ_u is given by Eq (2-13). It is assumed that $\xi(k)$ and $w(k)$ are uncorrelated and that $x(j)$ is uncorrelated with each $\xi(k)$ and $w(k)$ for $j \leq k$.

Letting

$$y(k) \triangleq \begin{bmatrix} z(1) \\ z(2) \\ \vdots \\ z(k) \end{bmatrix} \quad (3-12)$$

it follows from Theorem 2-1 that

$$\hat{x}(j|k) \triangleq E[\hat{x}(j)y(k)^T]E[y(k)y(k)^T]^{-1}y(k) \quad (3-13)$$

represents the best linear minimum variance estimate of all past state variables $\hat{x}(j)$ given the measurements $z(1), \dots, z(k)$. The j^{th} state estimate, $\hat{x}(j|k)$, is defined analogously. The relevant error covariance matrices are

$$P(j|k) \triangleq E[(\hat{x}(j|k) - x(j))(\hat{x}(j|k) - x(j))^T] \quad (3-14)$$

$$P(j|k) \triangleq E[(\hat{x}(j|k) - x(j))(\hat{x}(j|k) - x(j))^T] \quad (3-15)$$

From Eq. (3-5) and (3-13) it follows easily that

$$\hat{x}(k+1|k) = \Theta(k)\hat{x}(k|k) \quad (3-16)$$

If $\hat{x}(k+1|k)$ is known and a measurement of the form Eq. (3-6) is secured, it follows from the fundamental theorem of static updating of Gauss-Markov estimates [6] that

$$\hat{x}(k+1|k+1) = \hat{x}(k+1|k) + E[\hat{x}(k+1)\tilde{z}(k+1)^T]E[\tilde{z}(k+1)\tilde{z}(k+1)^T]^{-1}\tilde{z}(k+1) \quad (3-17)$$

where

$$\tilde{z}(k+1) \triangleq z(k+1) - \hat{x}(k+1|k) \quad (3-18)$$

Using arguments identical to those used in establishing Eq. (3-16) it is easy to establish that

$$\hat{z}(k+1|k) = H(k+1)\hat{x}(k+1|k) \quad (3-19)$$

From Eq. (3-19), and relations (3-6) and (3-18), it follows that

$$\tilde{z}(k+1) = H(k+1)\{x(k+1) - \hat{x}(k+1|k)\} + w(k+1) \quad (3-20)$$

From Eq. (3-20) it then follows that

$$E[\tilde{z}(k+1)\tilde{z}(k+1)^T] = H(k+1)P(k+1|k)H(k+1)^T + R(k+1) \quad (3-21)$$

where

$$R(k+1) \triangleq E\{w(k+1)w(k+1)^T\} \quad (3-22)$$

From the projection theorem for Hilbert space [6] it follows that $\hat{x}(k+1|k)$ and $x(k+1) - \hat{x}(k+1|k)$ are orthogonal, and so from Eq. (3-20)

$$\begin{aligned} E(x(k+1)\tilde{z}(k+1)^T) &= E\{x(k+1)(x(k+1) - \hat{x}(k+1|k))^T\}H(k+1)^T \\ &= E\{(x(k+1) - \hat{x}(k+1|k))(x(k+1) - \hat{x}(k+1|k))^T\}H(k+1)^T \\ &= P(k+1|k)H(k+1)^T \end{aligned} \quad (3-23)$$

From Eqs (3-23), (3-17), and (3-21) it follows that

$$\begin{aligned} \hat{x}(k+1|k+1) &= \hat{x}(k+1|k) + P(k+1|k)H(k+1)^T [H(k+1)P(k+1|k)H(k+1)^T \\ &\quad + R(k+1)]^{-1} [z(k+1) - H(k+1)\hat{x}(k+1|k)] \end{aligned} \quad (3-24)$$

Recalling Eqs. (3-8), (3-9), (3-10), (3-14), and (3-15) one can easily establish that

$$H(k+1)P(k+1|k)H(k+1)^T = H(k+1)P(k+1|k)H(k+1)^T \quad (3-25)$$

and

$$H(k+1)\hat{x}(k+1|k) = H(k+1)\hat{x}(k+1|k) \quad (3-26)$$

Defining

$$K(k+1) \triangleq P(k+1|k)H(k+1)^T [H(k+1)P(k+1|k)H(k+1)^T + R(k+1)]^{-1} \quad (3-27)$$

the state update equation can be put in the familiar form

$$\hat{\mathbf{x}}(k+1|k+1) = \hat{\mathbf{x}}(k+1|k) + \mathbf{K}(k+1)[\mathbf{z}(k+1) - \mathbf{H}(k+1)\hat{\mathbf{x}}(k+1|k)] \quad (3-28)$$

The covariance equations are derived in the usual fashion from Eqs. (3-16), and (3-28) to obtain

$$\mathbf{P}(k+1|k) = \mathbf{\Theta}(k)\mathbf{P}(k|k)\mathbf{\Theta}(k)^T + \mathbf{S}(k)^T\mathbf{Q}(k)\mathbf{S}(k) \quad (3-29)$$

where

$$\mathbf{Q}(k) = \mathbf{E}[\xi(k)\xi(k)^T]$$

$$\mathbf{P}(k+1|k+1) = [\mathbf{I}(k) - \mathbf{K}(k+1)\mathbf{H}(k+1)] \mathbf{P}(k+1|k) \quad (3-30)$$

and $\mathbf{I}(k)$ is an appropriately sized identity matrix.

The relations (3-16), (3-28), (3-29), and (3-30) are similar in form to the discrete Kalman filter equations, but are actually somewhat different in structure since, \mathbf{K} , $\mathbf{\Theta}$, and \mathbf{P} are all matrices whose dimension increases at each propagation step. Fortunately, the calculation required in Eq. (3-27) that computes a pseudoinverse is of constant size; however, this is not the case with $\Phi(k)$ in that $\Phi(k)^+$ must be calculated at each time step. The next section describes a recursive set of equations for computing the required pseudoinverse $\Phi(k)^+$.

4. RECURSIVE CALCULATION OF $\Phi(k)^+$

The problem of recursively calculating $\Phi(k)^+$ falls naturally into two cases, $\Phi(k)$ singular or non-singular. The non-singular case is handled easily by the following well-known theorem

Theorem 4-1 If $\mathbf{A} = \mathbf{A}^T$ has the form

$$\mathbf{A} = \begin{bmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{bmatrix} \quad (4-1)$$

and if \mathbf{A}_{11}^{-1} is known, then if \mathbf{A}^{-1} exists, it can be expressed in the form

$$\mathbf{A}^{-1} = \begin{bmatrix} \mathbf{J}_{11} & \mathbf{J}_{12} \\ \mathbf{J}_{21} & \mathbf{J}_{22} \end{bmatrix} \quad (4-2)$$

where

$$\begin{aligned} \mathbf{J}_{22} &= [\mathbf{A}_{22} - \mathbf{A}_{21} \mathbf{A}_{11}^{-1} \mathbf{A}_{12}]^{-1} \\ \mathbf{J}_{12} &= \mathbf{J}_{21}^T = -\mathbf{A}_{11}^{-1} \mathbf{A}_{12} \mathbf{J}_{22} \\ \mathbf{J}_{11} &= \mathbf{A}_{11}^{-1} + \mathbf{A}_{11}^{-1} \mathbf{A}_{12} \mathbf{J}_{22} \mathbf{A}_{21} \mathbf{A}_{11}^{-1} \end{aligned} \quad (4-3)$$

The theorem is proved by direct calculation.

The singular case can be treated by using Theorem 4-1 to obtain of the inverse of the maximal non-singular submatrix of \mathbf{A} and then using the algorithms developed below to calculate \mathbf{A}^+ . More specifically, suppose that \mathbf{A}_{11} is a maximal non-singular submatrix of \mathbf{A} , that \mathbf{A}_{11}^{-1} is known, and that the rows and

$$A = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} \quad (4-4)$$

Then

- $$A_{11}F = A_{12}, \quad A_{21}F = A_{22}$$

$$(2) \quad A_{22} = A_{21}^{-1} A_{11} A_{12}$$

(Hence J_{22} defined in Theorem 4-1 is zero.)

(3) If $E \stackrel{\Delta}{=} \begin{bmatrix} I \\ F^T \end{bmatrix}$, then

$$A \approx E A_{11} E^T$$

- (4) E has a left inverse L . In fact

$$L = (E_T E)_{-1} E_T$$

(5) $A'' = EL$

$$(6) \quad A_+ = L A_{-1} L$$

Proof (1) Since $\text{rank } (A) = \text{rank } (A_{11})$, it follows that the

columns in A represented by $\begin{bmatrix} A_{12} \\ A_{22} \end{bmatrix}$ are each linear combinations of the columns of $\begin{bmatrix} A_{11} \\ A_{21} \end{bmatrix}$. Thus, there exists an F such that

$$\begin{bmatrix} A_{11} \\ A_{21} \end{bmatrix} F = \begin{bmatrix} A_{12} \\ A_{22} \end{bmatrix}$$

and the result follows.

- (2) From part 1 it follows that F is given by

$$F = A_{11}^{-1} A_{12}.$$

VII9-10

Substituting this into the expression $A_{21}F = A_{22}$ the result is obtained.

(3) This is a simple calculation using the expression for F in (2) as well as the result in (2).

(4) E has full column rank, hence $E^T E$ is invertible. Result (4) follows at once.

(5) Let $P \triangleq EL$. Then

$$P^2 = (EL)(EL) = E(LE)L = EL = P$$

so P is idempotent hence a projection. Also

$$P^T = (EL)^T = L^T E^T = E(E^T E)^{-1} E^T = EL = P$$

so P is a perpendicular projection

Next, suppose $y \in \text{range}(A)$. Then there exists an x such that $y = Ax$. By (3)

$$y = EA_{11}E^T x$$

Thus

$$Py = ELEA_{11}E^T x = EEA_{11}E^T x$$

$$= EA_{11}E^T x = y$$

It thus follows that

$$\text{range}(A) \subset \text{range}(P)$$

Conversely, suppose $y \in \text{range}(P)$

Then $y = Py$. Let $x \triangleq L^T A_{11}^{-1} Ly$.

Then

$$Ax = EA_{11}E^T L^T A_{11}^{-1} Ly = EA_{11}(LE)^T A_{11}^{-1} Ly$$

$$= EA_{11}A_{11}^{-1} Ly = ELy = Py = y$$

and so

$$\text{range}(P) \subset \text{range}(A).$$

The conclusion to (5) follows

$$(8) \quad \text{Let } R \triangleq L^T A_{11}^{-1} L.$$

Then

$$RA = L^T A_{11}^{-1} LEA_{11}E^T = L^T A_{11}^{-1} A_{11} E^T$$

$$= L^T E^T = (EL)^T = (A'')^T = A''.$$

Also

$$RA'' = L^T A_{11}^{-1} L E L = L^T A_{11}^{-1} L = R$$

Since $A = A^T$, the equations

$$RA = A''$$

$$RA'' = R$$

imply that $A^+ = R$.

Theorem 4-2 is implemented in two ways. Assume A_{11} represents the maximal non-singular principal matrix in Φ . Letting

$$A \triangleq \begin{bmatrix} A_{11} & \hat{\psi}(k+1)^T \\ \hat{\psi}(k+1) & \Phi_{k+1,k+1} \end{bmatrix}$$

one calculates

$$\Phi_{k+1,k+1} \hat{\psi}(k+1) A_{11}^{-1} \hat{\psi}(k+1)^T. \quad (4-5)$$

(The caret (^) indicates that redundant information in $\psi(k+1)$ has been discarded and the resulting matrix compressed.) If the computation 4-1 is non-zero, it follows from Theorems 4-1 and 4-2 that A is invertible and one uses Theorem 4-1 to calculate A^{-1} . A is then the maximal non-singular principal matrix in $\Phi(k+1)$. If A is singular, then A_{11}^{-1} is the maximal non-singular principal matrix in $\Phi(k+1)$ and one simply retains A_{11} . In either case one can use Theorem 4-2 to calculate $\Phi(k+1)^+$.

It can be shown that if the field is such that singularities of Φ only occur when the track T contains repeated points, then the above simplifies and one need never calculate the pseudoinverse. Instead one need only retain the maximal non-singular principal part of Φ and perform simple averaging on the repeated estimates and the corresponding entries in \underline{z} .

5. FILTER SENSITIVITY

This section develops covariance equations used to evaluate the performance of linear recursive filters, for example, Kalman filters. The performance criterion is the root mean square (rms) value of the divergence between the generalized non-Markovian truth model, Eq. (3-5), and the state estimate from a Kalman filter. The equations necessary to compute the "time" history of the rms statistics resulting from processing simulated noise measurements in a Kalman filtering algorithm are presented in this section.

A filter using a Markov model has the form

$$\begin{aligned} \underline{x}^*(k+1) &\triangleq \begin{bmatrix} x^*(k+1) \\ u^*(k+1) \end{bmatrix} = \begin{bmatrix} \theta_{11}^* & \theta_{12}^* \\ 0 & \theta_{12}^* \end{bmatrix} \begin{bmatrix} x^*(k) \\ u^*(k) \end{bmatrix} + \begin{bmatrix} \xi_s(k) \\ \xi_u(k) \end{bmatrix} \\ &= \Theta^*(k) \underline{x}^*(k) + \xi^*(k), \end{aligned} \quad (5-1)$$

with a measurement model,

$$z(k) = H^*(k) \underline{x}^*(k) + w^*(k). \quad (5-2)$$

The * is used to denote the suboptimal filter model equations. The generalized non-Markovian truth model used for the comparison is summarized by Eq. (3-5) along with the associated measurement model, Eq. (3-6).

Often, the individual state variables within a Kalman filter have no physical significance. In order to compare them to physical quantities of interest, it is necessary to use linear combinations of these state variables. One interpretation of this is that the state variables within the system are used as a shaping filter, so that many state variables may be used to model a single physical quantity. It is this physical quantity that is the desired output of the shaping filter. The significance of this is that the divergence used for performance evaluation will be defined as the difference between physical quantities. The divergence, therefore, is defined as

$$v(k|j) \triangleq A\hat{x}^*(k|j) - Bx(k) \quad (5-3)$$

where the matrices A and B are used to make the two sets of systems variables, $\hat{x}^*(k)$ and $x(k)$ comparable.

Suppose at time k a measurement has been secured to estimate $\hat{x}^*(k|k)$ by a Kalman filter. The Kalman filter equations used to propagate and update the estimate with an external measurement $z(k+1)$, are

$$\hat{x}^*(k+1|k) = \Theta^*(k) \hat{x}^*(k|k) \quad (5-4)$$

$$\hat{x}^*(k+1|k+1) = \hat{x}^*(k+1|k) + K^* \{z(k+1) - \tilde{H}(k)A\hat{x}^*(k+1|k)\} \quad (5-5)$$

where $\Theta^*(k)$ denotes the dynamics matrix of a given discrete system, K^* is the Kalman gain matrix, and \tilde{H} is related to H^* via

$$H^*(k) = \tilde{H}(k)A, \quad (5-6)$$

$\tilde{H}(\cdot)$ being the measurement matrix defining the simulated physical measurements

The propagation equation is derived by substituting Eqs. (3-5) and (5-4) into Eq. (5-3)

$$\begin{aligned} v(k+1|k) &= A\hat{x}^*(k+1|k) - Bx(k+1) \\ &= A\Theta^*(k)\hat{x}^*(k|k) \\ &\quad - B S(k) \{ \Theta(k) \underline{x}(k) + S(k)^T \xi(k) \}. \end{aligned} \quad (5-7)$$

An explicit recursive representation of the divergence, $v(k|k)$, can be derived by adding and subtracting $v(k|k)$ on the right-hand side of Eq. (5-7),

$$\begin{aligned} v(k+1|k) &= v(k|k) - \{ A\hat{x}^*(k|k) - Bx(k) \} \\ &\quad + A\Theta^*(k)\hat{x}^*(k|k) \\ &\quad - B S(k) \Theta(k) \underline{x}(k) - B \xi(k) \\ &= v(k|k) + A \{ \Theta^*(k) - I_k \} \hat{x}^*(k|k) \\ &\quad + B S(k) \{ I(k) - \Theta(k) \} \underline{x}(k) - B \xi(k), \end{aligned} \quad (5-8)$$

where the matrices I_k and $I(k)$ denote fixed and variable size identity matrices respectively.

The update equation is derived by a similar procedure starting from the definition of the divergence after the $k+1$ propagation step, then substituting Eq. (5-5)

$$\begin{aligned} v(k+1|k+1) &= A \hat{x}^*(k+1|k+1) - B \hat{x}(k+1) \\ &= A \hat{x}^*(k+1|k) - K^* [z(k+1) - \tilde{H}(k) A \hat{x}^*(k+1|k)] \\ &\quad - B \hat{x}(k+1). \end{aligned}$$

This last equation can be rearranged as

$$\begin{aligned} v(k+1|k+1) &= [A \hat{x}^*(k+1|k) - B \hat{x}(k+1)] \\ &\quad + AK^* [z(k+1) - \tilde{H}(k) A \hat{x}^*(k+1|k)]. \end{aligned}$$

The first term on the right-hand side is a divergence. The second term can be simplified by rearranging the measurement equation associated with the truth model, Eq. (3-2), and

$$\tilde{H}(k) = H(k) B, \quad (5-9)$$

to obtain

$$v(k+1|k+1) = [I_k - AK^* \tilde{H}(k)] v(k+1|k) - AK^* w(k) \quad (5-10)$$

The rms value of the divergence is the quantity of interest in the performance evaluation problem. Therefore from Eqs. (3-5), as well as, Eqs. (5-5) and (5-10), (5-1) and (5-8), the augmented system can be formed to compute the second moment of the divergence as

$$D(k+1|k) \triangleq \begin{bmatrix} D_{11}(k+1|k) & D_{12}(k+1|k) & D_{13}(k+1|k) \\ D_{12}^T(k+1|k) & D_{22}(k+1|k) & D_{23}(k+1|k) \\ D_{13}^T(k+1|k) & D_{23}^T(k+1|k) & D_{33}(k+1|k) \end{bmatrix} \quad (5-11)$$

$$= \pi(k) D(k|k) \pi(k)^T + w(k) Q(k) w(k)^T \quad (5-12)$$

and

$$D(k+1|k+1) = \Lambda(k) D(k+1|k) \Lambda^T(k) + \tau(k) R(k) \tau(k)^T. \quad (5-13)$$

where

$$\begin{aligned} D_{11}(k|j) &\triangleq E\{v(k|j) v(k|j)^T\} \\ D_{12}(k|j) &\triangleq E\{v(k|j) \hat{x}^*(k|j)^T\} \\ D_{13}(k|j) &\triangleq E\{v(k|j) \hat{x}(k|j)^T\} \\ D_{22}(k|j) &\triangleq E\{\hat{x}^*(k|j) \hat{x}^*(k|j)^T\} \\ D_{23}(k|j) &\triangleq E\{\hat{x}^*(k|j) \hat{x}(k|j)^T\} \\ D_{33}(k|j) &\triangleq E\{\hat{x}(k|j) \hat{x}(k|j)^T\} \end{aligned} \quad (5-14)$$

$$\pi(k) \triangleq \begin{bmatrix} I_k & A[\theta^*(k) - I_k] & B S(k) [I(k) - \theta(k)] \\ 0 & \theta^*(k) & 0 \\ 0 & 0 & \theta(k) \end{bmatrix} \quad (5-15)$$

$$w(k) \triangleq \begin{bmatrix} -B \\ 0 \\ I(k) \end{bmatrix} \quad (5-16)$$

$$\Lambda(k) \triangleq \begin{bmatrix} I_k - AK^* \tilde{H}(k) & 0 & 0 \\ -K^* \tilde{H}(k) & I(k) & 0 \\ 0 & 0 & I(k) \end{bmatrix} \quad (5-17)$$

$$\tau(k) \triangleq \begin{bmatrix} -AK^* \\ K^* \\ 0 \end{bmatrix} \quad (5-18)$$

$$Q(k) = E\{\xi(k)\xi(k)^T\} \quad (5-19)$$

$$R(k) = E\{w(k)w(k)^T\} \quad (5-20)$$

The covariance Eqs. (5-12) and (5-13) derived in this section necessary to compute the "time" history of the rms statistics are implemented in the next section to illustrate filter sensitivity.

6 ILLUSTRATIVE EXAMPLE

The optimal filtering technique described in Section 3 and the divergence analysis described in Section 5 are used in this section to assess the expected performance of a suboptimal Kalman filter which is based on a particular design model. The simulation model corresponds to the error dynamics of an inertial navigation system (INS) influenced by vertical deflections. The results are presented to illustrate the effect of maneuver-induced error associated with a suboptimal filter design.

The suboptimal filter model is formulated by first considering the gravity error model which is characterized in along-track and cross-track coordinates [4]. These gravity errors drive the INS error model which, in this particular case, is described in north/east coordinates. The gravity-INS coupling is accomplished through a heading-dependent transformation. The reference (truth) model for the divergence analysis embodies state variables in north/east coordinates, as in the optimal filter analysis, thereby eliminating the need for any heading-dependent transformations. The state variables implemented for this example represent the Schuler dynamics of a complete INS. These state variables are listed in Table 6-1.

The results of the performance evaluation are presented in the form of two time histories (Fig. 6-1). The particular rms statistic portrayed is the north velocity error. The dotted curve in Fig. 6-1 presents the true rms error resulting from the use of a suboptimal filter to process discrete measurements from a gravimeter. The INS is mounted on a platform traveling along a straight line,

TABLE 6-1
REFERENCE INS STATE DESCRIPTION

STATE	FRAME (AXIS)	DESCRIPTION	UNITS
1		time	hr
2	E	INS latitude error	min
3	N	INS heading error	min
4	N	INS north velocity error	kt
5	N	INS east velocity error	kt
6	N	INS platform tilt about N	min
7	N	INS platform tilt about E	min
8		INS north damp, integ. output	min/hr
9		INS east damp, integ. output	min/hr

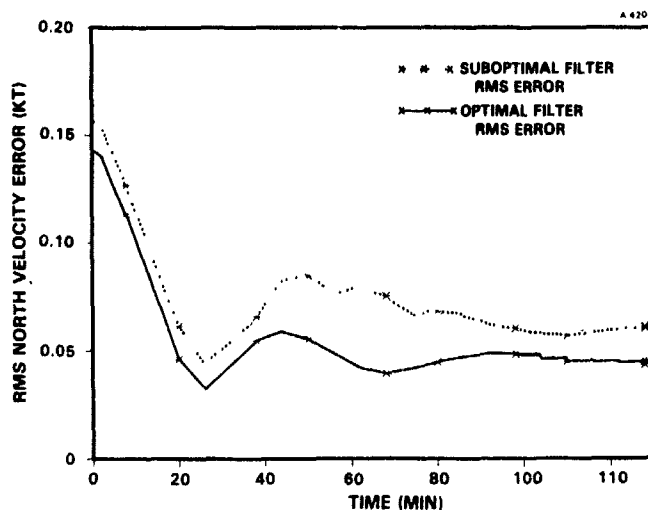


Figure 6-1 Representative Divergence Analysis Comparison

then beginning a circular maneuver (see Fig 6-2) The solid curve is the rms error predicted from optimal linear filtering

The accuracy of the suboptimal Kalman filter and the optimal filter converge to nearly identical values during the initial straight-line portion of the trajectory, indicating the adequacy of the suboptimal design for the straight-line portion. After the maneuver begins (at 20 min), the rms velocity errors associated with the two filters begin to diverge, demonstrating the filters' responses to the maneuver. The error associated with the suboptimal filter increases to nearly twice the pre-turn values before beginning to converge again. This response of the suboptimal filter is due in part to the coordinate transformation of the anomalous gravity statistics coupling with the INS. The suboptimal Kalman filter model assumes that the gravity field decorrelates in time and space as the vehicle moves along its trajectory. The truth model, however, assumes no such behavior and produces gravity statistics consistent with the underlying gravity field. Hence, the gravity statistics simulated in the Kalman filter and the truth model are different along the chosen curved path. This difference results in significant induced rms velocity error.

In summary this simulation demonstrates the effectiveness of the divergence analysis for determining the accuracy of a suboptimal filtering technique operat-

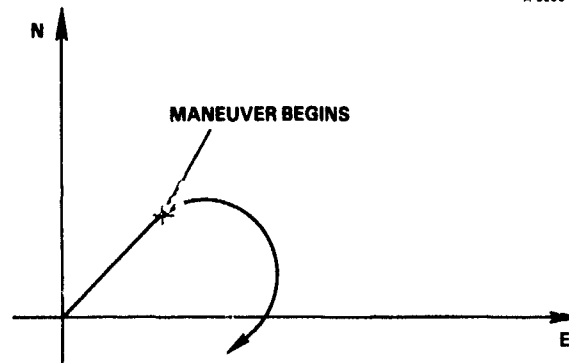


Figure 8-2 INS Platform Trajectory

ing in a spatial random field. When gravity measurements are processed by a suboptimal filtering technique, the results demonstrate the importance of matching the gravity correlations of the suboptimal filter and the truth model along the trajectory. When the filter models are "closely" matched (as along a straightline) then the accuracy is nearly optimal. Conversely, when there are large differences between the correlations (as along the curved trajectory), the accuracy degrades. The divergence analysis therefore becomes a design tool allowing the engineer to design suboptimal filters for spatial processes and access their expected accuracy.

REFERENCES

1. Edwards, R.M., "Gravity Model Evaluation for Precise Terrestrial Inertial Navigation: A System Accuracy Approach," Report AFIT/DS/EE/79-11, Thesis, 1979.
2. Foulis, D.J., "Relative Inverses in Baer *-semigroups," Michigan Math. J., Vol. 10, 1963, 85-84.
3. Levine, S.A. and Gelb, A., "Effect of Deflections of the Vertical on the Performance of a Terrestrial Inertial Navigation System," Proceedings of the AIAA Guidance, Control and Flight Dynamics Conference, p. 847, AIAA, New York, 1968.
4. Jordan, S.K., "Effects of Geodetic Uncertainties on a Damped Inertial Navigation System," IEEE Transactions on Aerospace and Electronic Systems, Vol. AES-9, No. 5, September 1973.
5. Kasper, J.F., Jr., "A Second-Order Markov Gravity Anomaly Model," Journal of Geophysical Research, Vol. 76, No. 32, November 1971.
6. Luenberger, D.G., Optimization by Vector Space Methods, Wiley, New York, 1969.
7. Nash, R.A., Jr., "Effect of Vertical Deflections and Ocean Currents on a Maneuvering Ship," IEEE Transactions on Aerospace and Electronic Systems, Vol. AES-4, No. 5, September 1968, pp. 719-727.
8. Rao, C.R., and Mitra, S.K., Generalized Inverse of Matrices and its Applications, Wiley, New York, 1971.
9. Willsky, A.S. and Sandell, N.R., Jr., "The Stochastic Analysis of Dynamic Systems Moving Through Random Fields," IEEE Transactions on Aerospace and Electronic Systems, Vol. AES-27, No. 4, August 1982.

PART VIII

Land Navigation Systems

A LOW COST INERTIAL/GPS INTEGRATED APPROACH TO LAND NAVIGATION

by

Mr D.G.Harris
GEC Avionics Limited, Guidance Systems Division,
Airport Works, Rochester, Kent ME1 2XX
United Kingdom

ABSTRACT

The positional accuracy obtainable from the Global Positioning System is much better than any preceding system and at a comparatively modest cost. This enables vehicles fitted with GPS receivers to be much more effective in operation and deployment. But to take full advantage of this accuracy, it is necessary that the navigation data is continuously available to the vehicle crew. While GPS is designed to be resistant to jamming it would be imprudent to assume that no significant periods of jamming will occur on the battlefield.

One solution which provides continuous data is to use a hybrid system combining GPS and a low performance Inertial Navigation System (INS). Data from these systems is combined in a statistical (kalman type) filter. This arrangement provides continuous data with no discontinuity of values, and redundancy to mitigate the effects of equipment failure.

This paper reviews the likely accuracy requirements for land navigation systems and ways in which these can be met with a low cost hybrid system. The results of the analysis of such a system, using test data from a low cost INS are included, demonstrating the effect of various GPS jamming patterns.

1. INTRODUCTION

In common with other military vehicles, the capability and hence complexity of army fighting vehicles is increasing, so that each unit is more effective. Factually there is a move to more mobile offensive operations which require flexibility in the ways that the vehicle can be used.

Inevitably these trends result in increased unit costs and hence smaller numbers of vehicles procured. This reinforces the need for flexibility of operation.

An important aspect of this is the ability to organise with high confidence, that groups or individual vehicles will be at a predetermined place at a given time. It may also be necessary that they travel along restricted corridors to the rendezvous.

To do this, each vehicle needs continuous knowledge of its position, route and direction of travel. A reliable navigation system will therefore be an important element of future generation military land vehicles.

Navigation equipment is currently classed as ancillary to the main function of a military vehicle and its cost will therefore be very closely scrutinised. There is a considerable challenge in meeting the demanding navigation requirements at a cost which is an acceptably small proportion of the total vehicle cost.

The navigation technique which comes closest to meeting the operational and performance requirements for fighting vehicles is Inertial Navigation (IN). A characteristic of all IN is that the navigation errors, position and velocity, grow steadily after the system is initialised. It is also necessary to know accurately the position of the vehicle at the point of initialisation. An INS with low error growth rates, say 0.5nm per hour, requires very stable and accurate gyroscopes and accelerometers; these are costly and such a unit would be priced around \$150,000.

It is possible to have most of the advantages of INS, but at much reduced cost, if an independent means for position fixing is included in the navigation system. GPS Navstar offers this prospect in the next decade.

This paper discusses the expected performance requirements for navigation systems and how well a hybrid of GPS and a low cost INS could meet these.

2. NAVIGATION SYSTEM REQUIREMENTS

The fundamental requirements for navigation systems are likely to remain unchanged until the year 2000. The operational requirements can be summarised as follows:

- (i) Continuously determine vehicle position and heading
- (ii) Display steering commands and distance-to-go to rendezvous points
- (iii) Rapid availability at start-up

- (iv) (Sensible) world wide operation
- (v) Minimum crew work-load
- (vi) Ability to accept and store pre-programmed positions and routes
- (vii) Strong preference for a non-radiating system

Typical accuracy requirements are:

Position	10 to 20 m CEP
Azimuth	1.5 mil rms
Elevation	1 mil rms

These applications again require a high quality full INS to meet all the performance criteria and it may still be necessary to use special techniques like zero-velocity updates to achieve the position accuracy.

2.1 Navigation Performance Requirements

These can be divided broadly into three accuracy bands dependant on the application.

2.1.1 Moderate Accuracy, Position Only

For vehicles required to navigate within visual distance of other units or landmarks over 10 or 20 km, an accuracy of 1 to 2% of distance travelled is likely to be acceptable.

An independent means for occasionally finding true position is necessary, for initiating the navigation system and updating en-route.

2.1.2 Accurate Stationary Position; Accurate Pointing

For fire support-team vehicles the choice of operating position is more dependant on cover and field of view, than the exact co-ordinates. However, once positioned, the crew need to know their position accurately in order to measure target location. This measurement is calculated from the range, bearing and elevation of the target relative to the observer.

A typical accuracy requirement is to locate the absolute position of a target to 25m at an observation range of up to 5 km.

Currently only the highest quality full inertial navigation systems (MAPS type) will meet this requirement

There are two axis INS-type units which provide the required accuracy in bearing and elevation when the vehicle is stationary, but they need to be aided in some way to meet the position accuracy.

2.1.3 Accurate Position Under All Conditions; Accurate Attitude When Stationary

All indirect fire weapons, especially long range rocket launchers and Self Propelled guns, need to know their position accurately throughout a mission. They also need high accuracy elevation and bearing of the launcher when the vehicle is stopped.

2.2 Navigation System Costs

The accuracy of all the outputs from a navigation system are basically dependant on the precision with which present position is measured. There are several techniques for doing this consistent with the military vehicle environment and they cover a wide range of accuracy and cost. The tendency is to prefer inertial and odometer based systems because of their complete autonomy, but the higher accuracy equipments are costly.

Radio location-based systems are much lower cost, but they are vulnerable to external interference and do not provide directly the useful secondary outputs (vehicle heading and elevation) of IN systems.

There is a wide range of performance and hence cost possible with INS, determined largely by the quality of gyroscope and accelerometer used. The range of system price varies from around \$35,000 for the medium accuracy odometer aided type to \$140,000 for the high accuracy full INS required for gun laying. This latter value is a large proportion of the cost of many military vehicles and therefore an unattractive solution. Much of the cost is associated with meeting the positional accuracy as opposed to attitude accuracy. The availability world wide of the Navstar GPS in the early 1990s will offer an alternative solution providing 3 dimensional positioning data to an accuracy of 15m SEP at a projected cost as low as \$6,000, per system. However, in common with all radio systems, GPS is vulnerable to jamming and it is therefore imprudent to propose it as a stand-alone system for battlefield use.

The availability in the future of GPS makes possible navigation systems which combine radio and INS techniques in a hybrid configuration. The advantage of this approach is high accuracy at a cost much less than INS of the same accuracy.

3. HYBRID RADIO/INERTIAL NAVIGATION SYSTEMS

The examples given in para. 2 for meeting the different performance requirements are all based on INS based equipments, as this represents present day thinking. Characteristics common to all these types of equipment when used alone, are that an initial period of alignment is required during which the vehicle must be stationary and the subsequent pattern of navigation errors is primarily a steady growth with time. Errors in position and heading, particularly, follow this pattern. Velocity and elevation errors tend to be oscillatory with periodicity of tens of minutes and therefore bounded in value, over a period of one or more hours.

To limit the position errors it is necessary intermittently to provide accurate position fixes, obtained independently of the INS. An alternative more complex method is to constrain the build-up of velocity errors using an on-board speed measurement or by regularly stopping for a few seconds (zero velocity updating).

The error patterns for radio navigation systems are normally not correlated with the time from mission start. Over a period of one hour and total movement of say 50Km the position errors are essentially random about a bias value.

Therefore the primary characteristics of the error patterns for INS and radio navigation equipments are quite different taken over even short time periods.

Because of the complementary nature of the error patterns for radio and IN type systems, they can be used in a mutually beneficial combination to enhance the performance of each, using a Kalman filter arrangement. There is the added benefit of redundancy should either sub-system fail and the removal of the restriction of several minutes stationary alignment of the INS before moving off. This latter benefit is due to the fact that most of the alignment period is required to assess the vehicle heading relative to true north. In a hybrid system it is possible to make the error in this assessment converge rapidly, while the vehicle is travelling normally.

If such a hybrid is considered for the land navigation requirements previously described, the choice of suitably accurate radio systems is very limited. The most obvious choice for the 1990s and beyond is GPS (1). The coverage will be world-wide with accuracy of 15m SEP, the user equipment is completely covert and it is expected that receivers will be made in very large numbers and hence low cost. It is also possible to derive speed and direction of travel from GPS signals so that a hybrid system has redundancy for all basic navigation functions.

3.1 Benefits of INS/Radio Hybrid System

The most useful way for combining the data from systems with complementary error characteristics is in a statistical filter of the Kalman type, a well established technique. The computing requirements are modest by comparison with that for either INS or GPS and can be accommodated in the processor of one of these sub-systems for a small additional cost of around \$100.

A major advantage of the hybrid/Kalman filter arrangement is that the variability of the INS performance can be much reduced on each journey, relative to the stand-alone values. This arises because the uncertainties in the IN sensor performance have two parts. The first is a random component which varies throughout a mission and the second is a value which varies from switch-on to switch-on, while staying sensibly constant during each mission. The second component is estimated by the Kalman filter during an operation and subsequently used to reduce the error build up in the INS. This assumes increased importance if the radio navigation data becomes suspect, as the hybrid system then reverts to unaided INS but with a reduced error growth pattern. This situation has obvious advantages where the radio reception may be only locally jammed during parts of an operation. This is seen as very likely for GPS in view of the precautions taken in the system design against general jamming.

The initial alignment of an INS takes several minutes if it has not been pre-aligned, and the vehicle must be stationary throughout the process. In a hybrid system it is possible to manually input the vehicle's approximate heading and move off immediately, using GPS data to rapidly reduce the heading error as the journey progresses. If the GPS also is starting 'from cold' it is possible to use the less accurate 'CA' code initially, which is rapidly acquired, until the full accuracy 'P' code data is processed.

3.2 Choice Of Hybrid System Components

The simplest form of GPS receiver is a single channel type which sequentially processes the data from the four most appropriate satellites and provides a position and velocity update every 5 s. This is very adequate for use in a hybrid filter. The estimated weight and power consumption for a mid 1990s equipment is 3 lb and 10 W.

The choice of INS is much wider and the essential factor is the quality of the sensors. High grade gyros and accelerometers have stable statistical errors but they are in high cost INS. Low grade sensors have a large cost advantage but may not be sufficiently stable in-run to provide acceptable stand alone INS performance during periods of loss of radio data.

To be effective in satisfying the cost/performance criteria a hybrid system must be based on a relatively cheap odometer aided inertial system which meets the generalised 1 to 2% requirement. GEC Avionics make such a system, GEC LNS (2). Fig. 1 is a picture of the Inertial Measuring Unit (IMU). This is based on strapdown INS design and uses 2 single degree-of-freedom floated gyros, type GI-G6, and 2 force-feedback accelerometers. The gyros have a run-to-run repeatability of a few °/h and an in-run value of 0.5°/h exponentially correlated with a time of 30 min. A novel spatial commutation mechanism enables these gyros to be used to gyrocompass within an error of 4/5 mil P.E. The weight and power consumption of the IMU is 15 lb and 30 W.

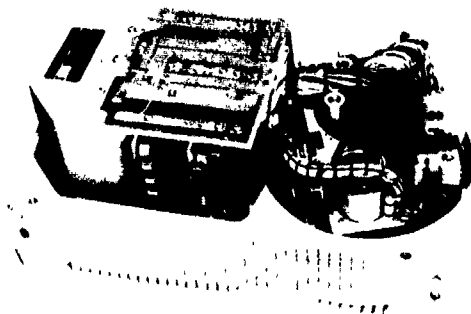


Fig. 1 GEC Inertial Measuring Unit

Trials of a stand alone system in a variety of vehicles, wheeled and tracked, light and heavy, have produced an accuracy of 1.5% CEP of distance travelled. The terrain used varied from rough hilly areas in military training grounds through curving narrow lanes to highways.

As a result of detailed records from these test runs, good statistical data is available on the position error growth and the variation of heading error with time. This database is therefore very suitable as a practical source for simulating the behaviour of the type of IN-based system likely to produce an economical hybrid system.

4. BASIS OF SIMULATION

The potential benefit from a hybrid system is to obtain high quality navigation performance from an equipment of modest cost. The approach chosen for the simulation exercise therefore, is to assume that the INS system and the GPS receiver are standard units with the simplest form of statistical filter to combine the two sets of data. A five state filter is used.

The errors modelled are:

- (i) Dead-Reckoning position east.
- (ii) Dead-Reckoning position north.
- (iii) Odometer scale factor.
- (iv) Heading angle.
- (v) Heading drift rate.

For GPS position the error is assumed to be of white noise form with an rms value of 20 metres. This is a little pessimistic compared with the limited test data currently available.

The position error statistics for the DR navigation system are extracted from a series of tests run over a 30 km course which involved sharply undulating terrain and many heading changes. This latter aspect is shown in Fig.2.

Statistics for heading and heading drift rate were extracted from the same series of tests, together with an estimate of odometer scale factor variation. The simulation program effectively runs the navigation system over the 30 km course mentioned earlier, but using various initial conditions for the INS and varying periods of interruption of the radio data, to simulate jamming. The outputs are radial position error and heading error.

5. SIMULATION RESULTS

As a datum, Fig. 3 shows the error growth on a base of distance travelled for the rms values of the set of unaided DR system trials. The second plot, Fig. 4, shows the effect of uninterrupted GPS position fixes at a frequency of one per 5 s. This represents the most accurate performance likely from the hybrid system, using a simple 5 state filter. Not unexpectedly, the position error is rapidly bounded to a value somewhat less than that for GPS alone. The more interesting result is the way that the heading error is bounded to a value around 4.5 mil.

There is no general agreement on the form that jamming of GPS might take, except that it is most likely to be in localised areas, because of the logistics of supplying and protecting effective jamming sources. To obtain a measure of this effect the following situation is assumed in the simulation.

The initial part of a mission is in an area free from jamming. The central part of the mission is subject to heavy jamming and signals are received for only one minute in each four minute period. The final one third of this mission is again in area with no jamming.

Fig. 5 shows the heading and position errors resulting from a transit through this jamming regime.

A further simulation of operational interest is the case of rapid start-up, when there is not time to establish the vehicle heading by gyrocompassing. In this case the vehicle commander inserts a best estimate of heading from compass or relative bearing and the system is immediately switched to the 'navigate' mode. For the simulation it is assumed that heading estimate has an rms error of 4° (71mil). Fig. 6 shows the resultant heading and radial position error for the same 30km route, with no jamming of GPS.

Fig. 7 shows the results of a combination of 4° initial heading error and the jamming environment used for Fig. 5.

Earlier simulation has shown that faster GPS update rates produce more rapid bounding of the heading error. The update rate used in all the simulations is one per 5s which is easily achieved with a single channel GPS receiver.

The simulation results are the square root of the variance values in the state matrix, that is the 1 σ values.

6. DISCUSSION OF HYBRID SYSTEM RESULTS

There are two aspects to the results. First the improvement in position accuracy which occurs whether the vehicle is moving or stationary and second the bounding of heading error, which requires that the vehicle moves. Hence the horizontal axis is distance travelled rather than time.

The general form of the results is not unexpected; the interest lies in the quality of IN system that is likely to yield, in a GPS/IN hybrid, a navigation performance with wide application. This brings the benefits of simpler logistics and lower initial price through larger quantity purchases.

The radial position error reduces to a little less than the GPS value in a very short time and the error does not increase greatly when the GPS updates becomes less frequent. In two axis IN-based heading reference units, the effective random walk component of heading error is an important parameter, determined jointly by the gyro and the heading algorithm. Simulation with values 50% greater than that measured

on the trials does not show a much degraded position error; the increase is less than 10%.

The heading error simulations normally assumed a 7 mil rms initial value i.e. the result of gyrocompassing. With updates from GPS every 5 s, the error is reduced to around 5.5 mil after 5 km and reduces to below 4.5 mil after 20 km. These values are sensitive to heading random-walk error in the Dead-Reckoning system. The value used is 22 MIL/vh. This can be reduced significantly by small changes to the IN mechanisation, and the resulting performance is then comfortably in the range defined for observation, reconnaissance or possibly re-supply vehicles (see 2.1.2). The most important logistic benefit is that the DR part of this hybrid is a natural candidate for the simpler navigation applications (see 2.1.1) on a cost and performance basis.

7. SUMMARY

Military vehicle fleets of the future are likely to be smaller than at present because unit costs will be forced up by the need for sophisticated protection and attack capability. In this situation, a key factor in obtaining greatest effectiveness from each vehicle is that they should be capable of maximum flexibility in the field. This is greatly helped by a reliable, self-contained navigation system on all prime vehicles. However, the cost of currently available equipment with the required accuracy, is high and it is fitted only when the need is absolutely essential. The widespread installation of tactically useful navigation systems is dependant on a significant reduction in the cost of reliably achieving a navigation and steering accuracy in the tens of metres range.

As GPS becomes generally available in the early 1990s, the use of GPS/INS hybrid navigation systems offers the prospect for reconciling the cost/performance requirements with low technical risk for a variety of types of vehicle. A system to meet the requirements of para 2.1.2 would cost around \$40,000. At the beginning of this period the most stringent requirements, for gun laying, will still need a high quality INS, but longer term projections suggest that even this requirement may be amenable to use of a lower cost hybrid system.

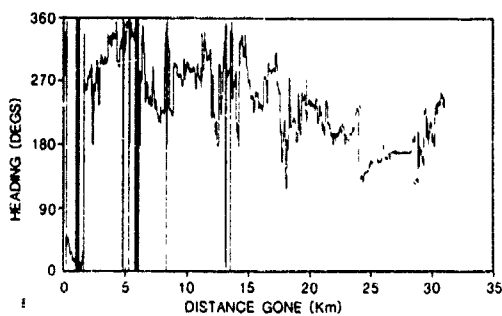


Fig. 2 Heading Variation Along Test Run

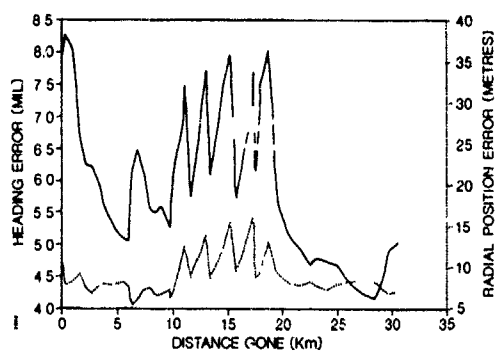


Fig. 5 Position and Heading Errors as a Function of Distance Travelled, Intermittent GPS During Mid Mission

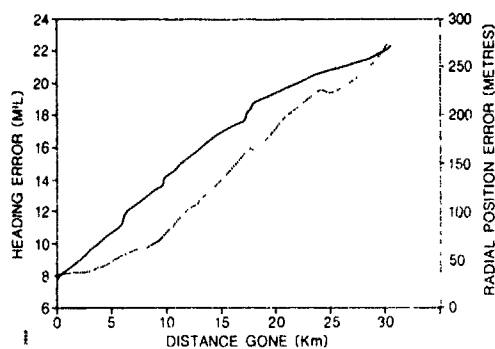


Fig. 3 Heading and Position Error Growth as a Function of Distance Travelled for Unaided Dead-Reckoning System

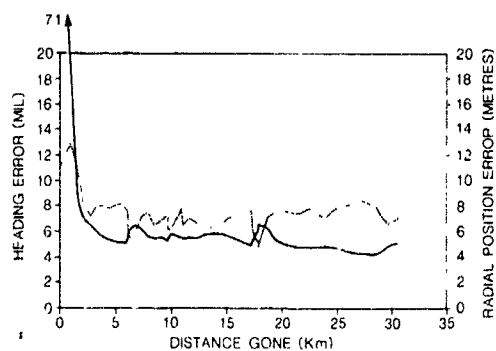


Fig. 6 Position and Heading Errors as a Function of Distance Travelled for Hybrid System with Continuous GPS, Starting with a Heading Error of 4° (71mil)

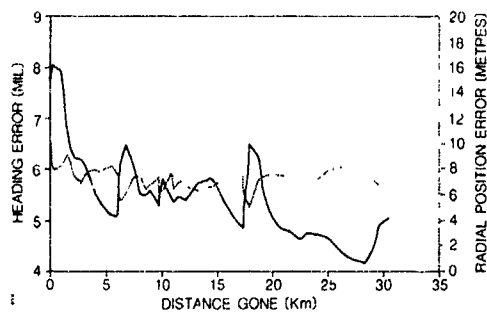


Fig. 4 Heading and Radial Position Error as a Function of Distance Travelled for Hybrid System with Continuous GPS Signal

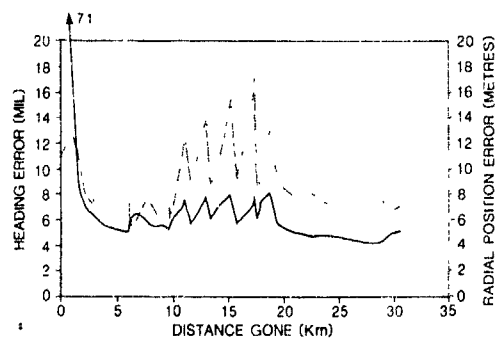


Fig. 7 Position and Heading Errors as a Function of Distance Travelled for Hybrid System with Intermittent GPS During Mid Mission, starting with a Heading Error of 4° (71mil)

8. REFERENCES

- 1 Principles Of Operation Of NAVSTAR and
System Characteristics
R.J. Milliken and C.J. Zoller.
Journal Of The Institute Of Navigation
Vol 25 No. 2 Summer 1978

GPS NAVSTAR System Overview
Col. J. Porter; P. Kruh; Sq. Ldr. B.
Sprosen.
NAVSTAR GPS Dept Of The Airforce USA.
- 2 An Affordable Land Navigation
System (LNS) For Large Scale Users.
D.G. Harris and J.T. Anders
Royal Institute of Navigation
'NAV 85' Conference, York, England.

A LAND NAVIGATION DEMONSTRATION VEHICLE WITH A COLOR MAP DISPLAY FOR TACTICAL USE*

by

E.J.Nava, E.E.Creel, J.R.Fellerhoff and S.D.Martinez
Sandia National Laboratories
Albuquerque, NM 87185
United States

ABSTRACT

A Land navigation Demonstration Vehicle (LDV) has been assembled which fully automates the navigation task and provides the operator with a color map display derived from Digital Terrain Elevation Data (DTED). The system relieves the operator of the burdens associated with the tactical use of paper maps by providing accurate 3-dimensional position information using a strapdown inertial navigation platform aided by the Sandia Inertial Terrain Aided Navigation algorithm (SITAN). The map display and navigation instruments consist of a multi-processor SANDia Aerospace Computer (SANDAC) and a commercial Image Processing System (IPS). These interactive devices allow real-time map annotation and corrections of vehicle position errors.

INTRODUCTION

In military land operations, the navigation task is presently done using topographic maps, a compass, and visual correlation with terrain features. In ideal conditions, manual navigation from terrain maps in unfamiliar territory is difficult, and map-determined positions are frequently erroneous. In tactical situations, this process is even less accurate because of high operator workload and limited visibility. In addition, vehicles such as tanks must operate 'buttoned up' so the field of view is restricted, further limiting navigational capability.

Automatic navigation equipment would be of great use to military land forces. A map display showing the local terrain would also be of benefit for advance mission planning, and especially for operations where visibility is limited (1). The map display could then be coupled to the navigation equipment and the vehicle's position presented on the map and adjusted as the vehicle moves. This arrangement would remove the navigation burden from the vehicle operator and provide accurate position information.

With a terrain display system there is a need for data to generate the displays. Digital Terrain Elevation Data (DTED) can be used for this purpose. The DTED required for generating the terrain display can also be used to improve navigation accuracy by using a terrain-aiding algorithm. Sandia has done extensive work in the terrain-aided navigation field using an algorithm called Sandia Inertial Terrain Aided Navigation (SITAN), which is used in conjunction with DTED (2-4).

The objective of the project described in this paper was to demonstrate a real-time implementation of an automatic terrain-aided navigation system with a digital map display using present day technology. Some of the equipment used was obtained from other Sandia and US Army projects. The remaining equipment was purchased commercially. The ensuing discussion describes the hardware and software components of an experimental Land navigation Demonstration Vehicle (LDV) used to implement the project objective. Included are details regarding the operational characteristics, display functions, system integration, software design, and navigation performance.

SYSTEM HARDWARE OVERVIEW

The system hardware is shown in Figure 1. At the heart of the system is a multi-processor computer developed by Sandia called the SANDia Aerospace Computer (SANDAC) (5). This computer is a Motorola 68000-based system with three processors. The SANDAC computer architecture shown in Figure 2 is configured so that one processor functions as an executive while the other two function as slaves. Each slave executes instruction codes within a local (zero wait state) RAM area. The executive processor executes instructions from global RAM. Any processor may access any SANDAC global memory location. An address bus arbitration circuit on each slave processor prevents concurrent memory accesses. The computer outputs include four serial ports configured as RS232-C outputs, and a 4k-word, 16-bit parallel interface.

*This work was supported by the US Army Avionics and Research Development Activity (AVRADA) under reimbursable contract.

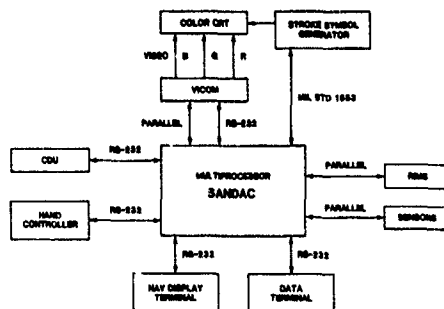


Figure 1. LDV System Hardware

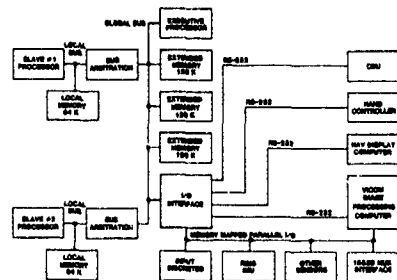


Figure 2. SANDAC Architecture

Interfaced to the SANDAC is an Inertial Measuring Unit (IMU), called Roll Isolated Measurement System (RIMS), developed at Sandia for spinning reentry vehicles. RIMS has two two-degree-of-freedom tuned rotor gyroscopes and three accelerometers. It is gimbaled in one axis and mounted so that the sensors can be rotated in a horizontal plane during gyrocompassing. During navigation, however, the gimbal is locked and the IMU is run in a pure strapdown mode. The gyroscope and accelerometer measurements are read by SANDAC. All navigation calculations are performed in a SANDAC slave processor. Other navigation sensors interfaced to SANDAC include a Paroscientific barometer and a Vascar odometer which measure altitude and distance traveled respectively and are also used for stabilizing the Inertial Navigation System (INS) mechanization.

SANDAC operation is controlled by the data terminal. This function is implemented using a COMPAQ portable computer operating as a smart terminal. This terminal is used to load software and data into SANDAC and to control software execution. A Maypro portable computer is also interfaced with SANDAC. It operates in a receive only mode and is used to provide a real-time display of navigation parameters.

The map display functions are selected by an operator through the Control Display Unit (CDU) and the Automatic Hand Controller (AHC). The Harris CDU is interfaced to SANDAC using a RS232-C serial interface. It has a CRT screen on which a display selection menu is presented. The AHC consists of a force proportional joy-stick and three discrete input switches which are used to control the vehicle cursor during operator input operations. The map displays are generated by a Vicon Image Processing System (IPS) running under a Firmware Operating System (FOS). The IPS was hardened by Sandia for vehicular use. The IPS is interfaced to SANDAC through both a serial and parallel interface. The serial interface is used to issue display commands, whereas the parallel interface is used for DTED transfers from SANDAC. The FOS software includes custom software which decodes input commands and implements the more advanced functions. The map displays are presented on a Rockwell/Collins color display system. The display system includes a stroke symbol generator which is interfaced to SANDAC through a MIL STD 1553B MUX bus.

The equipment was mounted in a prototype of a US Army High Mobility Multipurpose Wheeled Vehicle (HMMV), which was equipped with an AC power generator and related power conversion equipment. The equipment was mounted in 19 in racks when possible. Vibration isolators were used for all of the commercial quality equipment.

LAND DEMONSTRATION VEHICLE FUNCTIONALITY

INS Navigator

The RIMS IMU is the basis of the LDV strapdown INS mechanization, which for this implementation, exhibited an accuracy of between three and four percent of distance traveled. The RIMS sensor data are read by SANDAC and used to drive the navigation equations which are divided into attitude and position sections. The net iteration rate is 16 Hz for all navigation equations.

The vertical channel of the INS is damped using a second-order damping filter in conjunction with the barometer-derived elevation. The horizontal channel velocities are also damped by using second-order filters in conjunction with the odometer-measured velocity. The horizontal damping tends to suppress the oscillatory errors which are characteristic of an INS. The INS operates independently of SITAN.

During navigation, SITAN and display software are activated by the navigator software once per second. The start-up is phased so that the navigation display software does not read the variables during the time periods of SITAN or navigator execution.

LDV/SITAN

SITAN is an extended Kalman filter algorithm which recursively produces accurate position corrections to the vehicle reference trajectory from elevation measurements and DTED. The extended Kalman filter processes residuals forced from elevation measurements and predicted elevation measurements to recursively estimate errors in the reference trajectory. Elevation measurements are obtained from the navigator vertical position. Predicted elevation is obtained from plane-fit DTED corresponding to the area of vehicle operation. The SITAN functional configuration is shown in Figure 3.

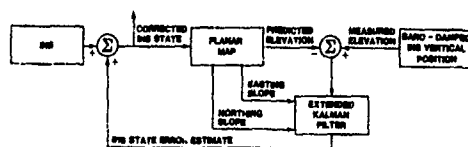


Figure 3. LDV/SITAN Implementation

Typical unaided inertial systems produce reference trajectories with large error that is unbounded in time. A SITAN-aided trajectory has small error bounded in time, with magnitude determined by DTED roughness and accuracy. Radial horizontal position error is typically less than 100 m CEP.

The SITAN algorithm utilized in the LDV system is based on the AFTI/SITAN algorithm (4). The AFTI/SITAN algorithm was designed specifically for use on low-flying, high-performance attack aircraft; specifically the AFTI/F-16. The SITAN algorithm design for the LDV system (LDV/SITAN) contains the operating modes and control logic of AFTI/SITAN but has been modified for use on a low-speed, low-dynamic land vehicle.

LDV/SITAN Design

The navigator error model utilized in the LDV/SITAN design consists of five states: three navigator position errors, navigator heading error, and odometer scale factor error. Both heading and scale factor errors are modeled as a random constant plus random walk. The random walk on heading error accounts for the fact that the heading error will change with time and direction of vehicle travel. The random walk on scale factor error accounts for scale factor error change due to vehicle tire slippage. The navigator vertical position error is also modeled as random constant plus random walk, where the random walk accounts for barometer error changes due to vehicle velocity, wind, and atmospheric changes. The INS horizontal errors are modeled as random constants.

The DTED utilized in the system is a 157 x 83 post planar map corresponding to a 15.6 x 8.2-km area near Edgewood, NM. This DTED was manufactured by Koogle and Poulos, an Albuquerque engineering firm. The DTED stored in SANDAC global RAM during operation is static, uncompressed, and geographically contiguous.

LDV/SITAN consists of an Acquisition mode, a Track mode, and the necessary logic to switch between the two (Mode Control Logic). In the Acquisition Mode of LDV/SITAN, a set of extended Kalman filters are executed in parallel to reliably determine the vehicle's position within a large uncertainty region. The parallel filters are configured spatially such that they cover a specified initial uncertainty region. Periodic checks are made to determine if any one filter can be identified as being close to the true vehicle position. LDV/SITAN consists of 21, 5-state, parallel Kalman filters spaced 400 m apart in a circular grid. All 21 filters utilize a nine-point plane fit, iterate at 1 Hz, and update after every 100 m of distance traveled. This acquisition configuration allows for a 500 m CEP initial vehicle position error. Since LDV/SITAN produces no position estimates during acquisition, it is desirable to minimize the distance traveled during acquisition. A small parallel filter spacing is used to promote fast covariance draw-down. Frequent application of the acquisition-to-track tests shorten acquisition

distance.

The LDV/SITAN Track Mode uses a single 5-state filter to continuously estimate the position of the vehicle as accurately as possible. The track filter iterates at 1 Hz and updates after every 100 m of distance traveled. The LDV track filter uses a nine-point plane fit and conventional covariance update.

Control Display Unit

The CDU is a multi-function component which provides the operator the capability to control and exploit the built-in LDV functions. The versatile display-control concept designed by Anacapa Sciences (6) consists of a 12-line (16 characters each) monochrome CRT display and a lower panel with several different types of controls. Line select keys are positioned adjacent to the 10 central lines of the display. The top line is reserved for user advisory information. The bottom line is used as a scratch pad area for echoing keypad entries. Prompting messages are displayed beside each line select key, indicating its particular function. These messages change according to the type of transaction in progress. Controls on the lower panel allow the user to enter numeric values into the system, jump back immediately to the primary menu display, and initiate a host of other special purpose functions. Design of the CDU dialog, which controls the color map presentation, was provided by Anacapa Sciences (6).

Upon system startup, the user is presented with a main menu display which contains various options for selection of the color map display presentation, including: change scale to 3, 6, or 12 km; alter contour line interval; turn contour lines on or off; select a relief shading or an elevation guide display; update the current vehicle location; enter reference and designated position symbols; orient the map in one of four cardinal directions; and generate intervisibility calculations at a user-selected location. Implementation details regarding these functions will be discussed in detail in the image processor software description. Also contained in the main menu display are additional control and auxiliary functions which are used to initialize the system and to control the SITAN navigation software.

Image Processing System

The Vicom FOS consists of an MC68000-based microcomputer interfaced to a number of general purpose, high-speed image processing modules. Figure 4 illustrates the five Vicom image processing modules. The Display Controller module refreshes the RGB monitor. The Image Memory module allows the storage of 16, 512 x 512 pixel images with 16 bits per pixel. This image memory is used for data base storage, as an image scratch pad area, and as a buffer for terrain displays. The Array Processor Module is used for computing 3 x 3 kernel convolutions at video frame rates and for image operations requiring a multiplier/accumulator. The Point Processor module is used for image operations which require ALU functions or a Look-Up Table (LUT). The Pipeline Controller is loaded by the microprocessor with microcode destined for the other modules in the system. This module is responsible for synchronizing all other image processing modules.

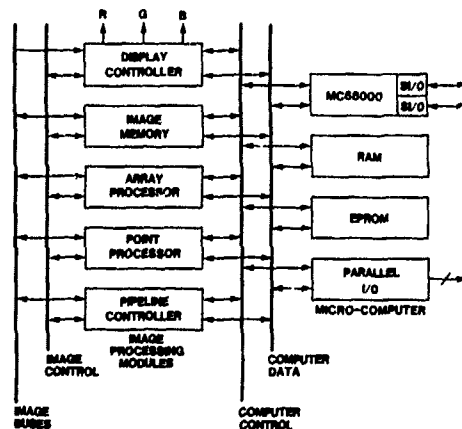
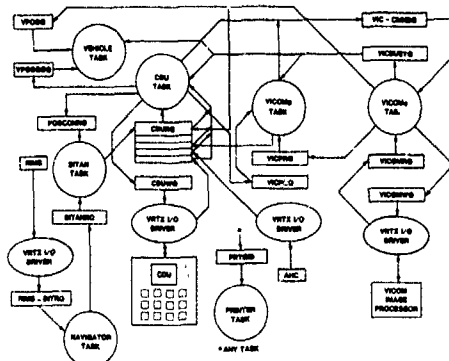


Figure 4. Vicom Image Processing Modules

LDV System Software

All application code for LDV was written in C and MC68000 assembly language. The primary software development environment consisted of an Alcyon/68000 C cross-compiler, assembler, librarian, and linker, running on a VAX 11/750 computer (7). Some of the IPS algorithms were simulated using the VAX native C compiler. Other IPS algorithms were designed and tested solely on the Vicom FOS due to their need for interaction with the high-speed image processing hardware.

Concurrent task management and intertask communication was effected using a Hunter & Ready Inc. VRTX/68000 real-time operating system, installed on each of the SANDAC processors. A set of extensions to the VRTX kernel was written to allow remote message queuing between tasks executing on different processors. Each task was assigned a priority, allowing the more time-critical tasks to execute ahead of the others. Figure 5 illustrates how the concurrent LDV tasks and the associated message queues were employed to interface and synchronize a portion of the system peripherals. Aside from the Navigator and SITAN tasks, all software elements depicted were implemented on the Executive processor. A brief description of each software task is presented below.



Operating systems were required to interface the VRTX kernels into the SANDAC hardware environment. Two versions were written, one for the slave processors and a second for the executive. Custom input/output drivers for communication with the various system peripherals were included in the executive processor operating system.

The FOS, on booting from EPROM, comes up in an interactive shell which allows the user to manipulate images by entering three letter command mnemonics followed by a number of parameters. This serial bit stream, flowing from the serial communications port to the shell program, was intercepted by a custom task which monitored the bit stream for a special character representing a header for custom functions. On detection of this special character, the bit stream, until termination, was then buffered and passed to the custom LDV software. In this way, it was possible for a user-defined software module to gain control of all of the resources of the IPS.

SANDAC Executive Processor

The system console Printer Task was written to synchronize the output messages from the various tasks to the console. A task performs output by first formatting a message into a static character buffer and then posting a pointer to this buffer to the printer task. The printer task uses the pointer to access the buffered message and then outputs the buffered message to the console.

The Clock Task maintains a real-time user clock (time-of-day) on the CDU main auxiliary page, based on the RIMS interrupt. This clock is initialized by the user during system startup.

The Vehicle Task executes at a 1-Hz rate, checking the SITAN estimated vehicle position for changes in east or north position greater than the pixel resolution of the map display (typically 100 meters), and issues commands to the image processor to relocate the vehicle position symbol accordingly.

The navigation display software implements two functions. It drives the 1553 interface to provide present position coordinates on the stroke symbol generator once a second, and it also provides a data stream of navigation parameters on the serial port for real-time display on the Kaypro computer every two seconds.

SANDAC Slave Processors

Inertial Navigator

Since the inertial navigation software is computationally intensive, it alone resides in slave processor 1. The software is coded in C and uses Motorola Fast Floating Point numeric representation for most variables. The floating point numbers are represented by 24 mantissa bits, 7 exponent bits, and 1 sign bit. This single precision floating point format is not accurate enough for an airborne INS implementation. Here the numerical imprecision effects are suppressed by the damping filters.

The gain tables for the alignment Kalman filter are included in the INS software. These values were coded in assembly language format and linked to the C code with external label references.

SITAN

The LDV/SITAN algorithm is coded in C and MC68000 assembly language and executes in Slave Processor 2. The Kalman filter and Mode Control Logic calculations utilize single-precision (32-bit) floating-point arithmetic. The nine-point plane fit calculations utilize single-precision (16-bit) fixed-point arithmetic. As shown in Figure 6, the slave processor executes the SITAN algorithm and geographic-to-state plane coordinate conversions at a 1 Hz iteration rate. Both the SITAN and coordinate conversion routines execute as a single task under control of the slave's VRTX operating system. The DTED is stored in global RAM and is accessed directly by the SITAN task. DTED access rates are 378 bytes/sec in acquisition mode and 18 bytes/sec in track mode.

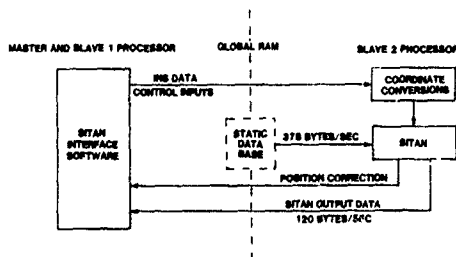


Figure 6. SITAN Software Implementation

The SITAN algorithm is well suited for implementation in a parallel processing environment, as it is computationally intensive but requires a low iteration rate and a small amount of input and output.

Image Processing System

The IPS software was implemented in a layered modular fashion. Referring to Figure 7, the software system consists of four levels. The Console/Protocol Handler, is responsible for maintaining communication between the SANDAC controller and the IPS. This handler verifies that the incoming messages satisfy a predetermined communication protocol. If the incoming message is incorrect, a retransmit is requested. All outgoing messages are also formatted with the proper header and trailer blocks by this module.

At the second level, the incoming commands are parsed, and the proper sequence of Function Modules are invoked to satisfy the requested command. If the incoming command is invalid, the host is notified of the error, and the command sequence is aborted. If a given command fails to complete successfully, recovery is attempted, and the host is notified of the unsuccessful command completion. At the third level, the Functional Modules are implemented. These will be described in more detail below. At the lowest level, various support functions are implemented. These generally maintain track of resources and access image memory or low-level IPS functions via the standard interpreter.

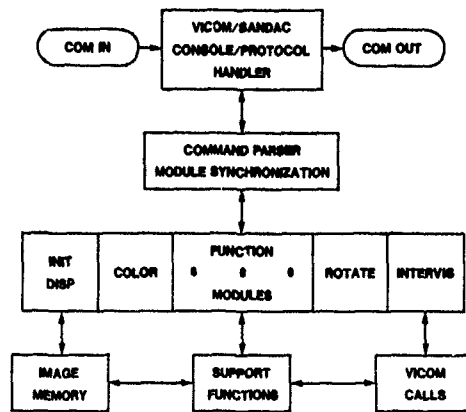


Figure 7. IPS Software Organization

Descriptions of the IPS function modules follow:

1. Initialize Global Variables - All IPS global variables are initialized.
2. Initialize Display - resets the display only. This function is often used when changing from one major display mode to another.
3. Color - This function is used to interactively adjust the saturation of the 32 available colors to compensate for variations between display monitors.
4. Symbols - This function paints, erases, and keeps track of five independent symbols. The symbols represent the immediate cursor, the vehicle, a designated position, a reference position, and an intervisibility reference position.
5. Coordinate Transforms - This function maintains knowledge of the present state of the IPS and appropriately transforms coordinate references to the internal coordinate system.
6. Shading Algorithms - Two types of terrain shading are available on the IPS. Slope shading is performed by convolving the terrain elevation data with a 3x3 kernel which approximates the cosine of the angle between the terrain normal and the incident radiation. (See Figure 8.) Elevation shading is performed by remapping the terrain data into a maximum of 16 equally spaced, user defined, quantization levels. The results of both elevation shading techniques modulate the saturation of the terrain color as specified by the cultural features encoded in the data base.
7. Contouring - Contouring is performed by quantizing the elevation data at the specified contour interval and then performing an edge detection followed by a threshold to locate the contours. These contour lines are then overlayed onto the displayed terrain.
8. Rotation - The rotation function rotates the data base such that the terrain display can have a north, south, east, or west up orientation. All references to map coordinates are also transformed as necessary.
9. Intervisibility - This function calculates the visibility of points within a radius of 120 pixels of the reference point.

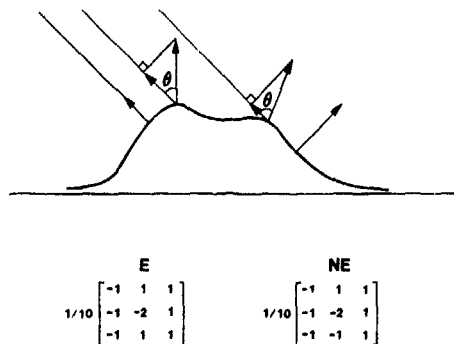


Figure 8. Terrain Slope Shading

The intervisibility calculation (see Figure 9) is done a ray at a time with a one-half-degree increment between rays until 45 degrees is reached. This process is repeated for the other seven octants. Along each ray, 120 incremental steps are taken. The increments are precalculated for the first octant and are stored in a table. Computationally efficient sign reversals are used to derive the incremental steps for the other seven octants. As steps are accumulated along a given ray, the slope of the Line Of Sight (LOS) is maintained. The present LOS is extrapolated to the next increment on the ray, and the extrapolated elevation is then compared with the elevation at the closest grid point. If the extrapolated elevation is less than the closest terrain elevation, the point is marked as visible, and the slope of the LOS is updated to reflect the higher horizon. If the extrapolated elevation is greater than the closest terrain elevation, no action is taken. This indicates the terrain point is not visible. The user can optionally specify an average tree height which biases every terrain elevation point in the calculation region excluding the observer point.

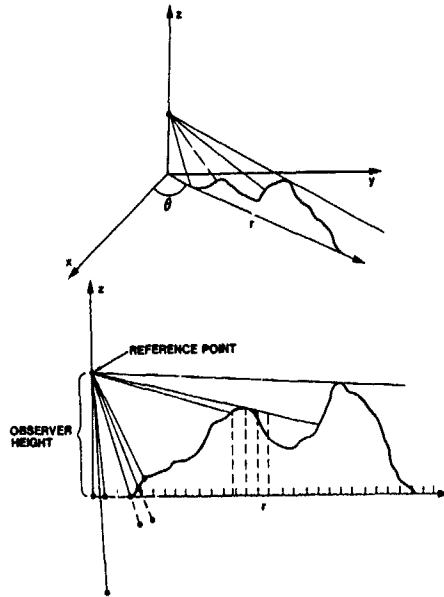


Figure 9. Intervisibility Calculation

NAVIGATION PERFORMANCE

The vehicle navigation system was tested at an area near Edgewood N.M., and the performance is illustrated in Figure 10. The road loop is approximately 30 km long and has eight landmarks. The navigation performance cannot be determined between the landmarks because the true vehicle location is not known, but navigation performance can be inferred from the fact that the vehicle was traveling on the road and the road loop has been accurately surveyed.

The system test started at landmark 3 with LDV/SITAN in automatic acquisition mode. For this particular test, acquisition occurred (denoted by * in Figure 10) after 56 updates, or 5.5 km of distance traveled.

LDV/SITAN remained in automatic track mode the remainder of the loop. The unaided navigator trajectory contains errors of several hundred meters, while the SITAN-aided trajectory radial error is generally less than 100 m. Table 1 lists the radial error for both the aided and unaided trajectories at each landmark along the road loop.

Table 1.
Aided and Unaided Radial Error

landmark	Navigator Error (m)	
	unaided	SITAN-aided
3	0	0
2	240	56
9	299	59
8	428	47
7	401	15
6	387	35
5	330	32
4	294	101
3	256	112

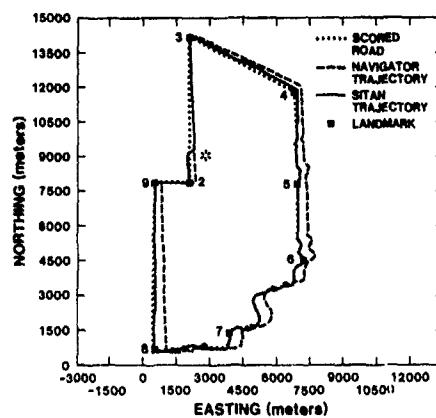


Figure 10. SITAN Performance at Edgewood

CONCLUSIONS AND SUMMARY

The LDV digital map display coupled with an automatic navigation system provides vehicle operators with capabilities beyond those of using paper maps with manual navigation. The computational capabilities can be exploited to give information such as continuous range, bearing, and intervisibility easily and accurately. In addition, the DTED used for the map display can also be used for accurate navigation.

An automatic navigation system coupled with a digital map display is feasible using current technology. Some of the demonstration equipment had excess capabilities and high-power requirements because it was designed for lab environments. A system designed exclusively for a military vehicle environment could be miniaturized and simplified considerably using new integrated circuit technology.

Moving map terrain-aided systems of this type will appear in military land vehicles as the necessary accurate DTED and very low-cost computers become available.

References

- (1) N. K. Shupe, "The Night Navigation and Pilotage System", in Amer Helicop Soc Conf Proc, May 1981, pp 18.1-18.23.
- (2) L. D. Hostettler, "Optimal Terrain-Aided Navigation Systems", SAND78-0874. Albuquerque: Sandia National Laboratories, August 1978.
- (3) E. J. Nava, "A Terrain-Aided Land Navigation System", SAND83-0470. Albuquerque: Sandia National Laboratories, June 1983.
- (4) D. D. Boozer et al, "SITAN Design for Low-Level Attack Aircraft", SAND84-2059, Vols I,III. Albuquerque: Sandia National Laboratories, May 1985.
- (5) C. R. Borgman and P. E. Pierce, "A Hardware/Software System for Advanced Development Guidance and Control Experiments", in AIAA Comput in Aerosp Conf Proc, October 1983.
- (6) S. P. Rogers, "The Integrated Mission Planning Station - Functional Requirements, Aviator-Computer Dialogue, and Human Engineering Design Criteria", Santa Barbara, Ca: Anacapa Sciences, August 1983.
- (7) "C68VX Reference Manual", San Diego, Ca: Alcyon Corp., June, 1984.

SATELLITE NAVIGATION SYSTEMS FOR LAND VEHICLES

by

Ronald A. Dork and Oliver T. McCarter
General Motors Advanced Engineering Staff
General Motors Technical Center
30200 Mound Road
Warren, MI 48090-9010
United States

ABSTRACT

Today's motorists often must confirm their route by referring to a roadmap. Looking at a map while driving creates a poor traffic situation, one even more acute for drivers of emergency service vehicles. To address this problem, development engineers at the General Motors Technical Center and Delco Electronics integrated an experimental GPS receiver into a GM Buick Park Avenue. The receiver is designed to acquire and sequentially track signals from four satellites. The vehicle's precise latitude, longitude, and altitude are determined and presented on a color cathode ray tube (CRT) map display in the car's instrument panel.

INTRODUCTION

This paper describes the development of an experimental application using a GPS receiver designed for ultimate use on automobiles and other land vehicles. Though mainly intended for U.S. defense purposes, the fact that the U.S. Government is sponsoring development of the satellite infra-structure represents a commercial spin-off opportunity for any private organization capable of developing low cost user equipment.

The Navstar satellites revolve around the earth at an altitude of 10,898 nautical miles and transmit precise and continuous navigation signals to any number of users over the entire globe.

This altitude corresponds to a 12-hour orbital period, a feature which enables the satellite's data to be up-loaded daily from earth stations located on American soil. In addition the high altitude achieves worldwide continuous 24-hour coverage with minimum signal distortion. Military tests conducted to date have confirmed the unprecedented accuracy and performance that GPS offers those having an application for its use (1).

A prototype GPS receiver designed to acquire and sequentially track signals from four satellites is installed in a GM Buick Park Avenue and a GMC van. By means of geometric triangulation, and a means of nullifying the user clock bias (which directly translates to user position error when multiplied by the velocity of light), the vehicle's precise latitude, longitude, and altitude are determined and presented on a color CRT map display located in the vehicle instrument panel.

As the vehicle moves along the streets, the CRT screen indicates the car's location on the map display as shown in Figure 1. As the vehicle travels along the map's secondary roads, marked in yellow, or onto major freeways, marked in purple, the car appears on the screen as a small flashing rectangle leaving a blue trail. The edges of waterways such as the Detroit River and Great Lakes are outlined in blue. The boundaries of different states are multi-colored.

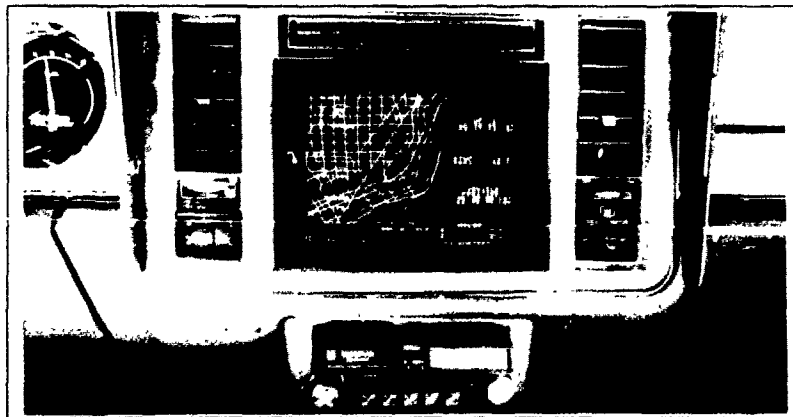


Figure 1 - MAP DISPLAY

A number of discrete zoom levels have been added to provide further detail when navigating in densely populated residential areas or city streets. A trip planning feature enhances operation by permitting the user to pre-program destinations which are also highlighted on the map display.

To evaluate the systems, test bed vehicles were built. They include a luxury sedan for user demonstrations and a mobile test van for quantitative data gathering, outfitted with a GPS receiver and an integrated display subsystem.

Initially, an overview of the Global Positioning System operation is described, followed by a functional description of the receiver for different application scenarios. The goals are addressed in light of the vehicle operational environment and from the perspective of the vehicle operator. This is followed by a description of the major components in the integrated system, concluding with vehicle testing, evaluation and test results.

GLOBAL POSITIONING SYSTEM

GPS is well suited for this task because it provides an indefinite number of users with accurate, continuous, worldwide, all weather, 24-hour coverage. The GPS system is organized into three main segments: the Space Segment consisting of the satellites transmitting coded messages of time and their orbital position to user receivers on earth, the Control Segment consisting of ground-based monitor and control stations to assure message integrity from each satellite on a daily basis, and the User Segment consisting of the indefinite number of passive receivers to receive the specially coded satellite signals.

The current constellation design requires a minimum of 18 Navstar satellites revolving about the earth when the system is fully deployed. They will be configured in six orbits so that any user set will be able to view at least four satellites simultaneously anywhere on earth at any time of the day or night. Because the orbital period is 12 hours, some satellites pass out of sight below the horizon as others rise to take their place. This enables the user to maintain continuous contact with at least four members of the constellation. During this development phase, the U.S. Government is maintaining seven active operational prototype satellites in orbit. The satellites are used to supply signals for the development of user equipment until the production satellites are launched.

The Control Segment consists of four earth-based monitor stations at precisely known coordinates to receive and monitor the data stream transmitted from each satellite as it passes over view. A comparison is made between the precisely known location of each monitor station and that computed as a result of receiving the satellite data.

The Control Segment also synchronizes the satellite's atomic clocks to standard GPS time. Computers then process these readings to determine the satellite's position, velocity, and clock error.

In essence, closed-loop control of each satellite's data base is maintained as errors in the satellite's broadcasted message are compensated and new data up-loaded to each satellite via a microwave data link as shown in Figure 2. The monitor stations will be based on American soil. Monitor stations are currently operated on Kwajalein Island and Hawaii in the South Pacific; Colorado Springs, Colorado; Ascension Island off the coast of Africa; and Diego Garcia Island in the Indian Ocean. The master control station which refreshes the satellite's data base is located at Falcon Air Force Base in Colorado Springs, Colorado.

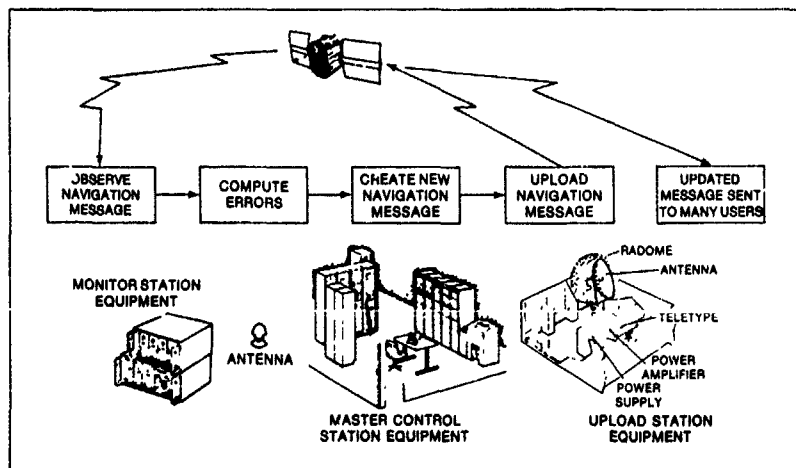


Figure 2 - CONTROL SEGMENT

Satellite Locations

The satellite positions must be known at least as accurately as the desired accuracy of the user's position. The coefficients in the satellite's almanac data block describe the mean orbits for each of the satellites. These coefficients are used by the receiver on a cold-start in order to determine which satellites to acquire. Once a satellite is acquired, the coefficients for its ephemeris are read by the receiver. The ephemeris describes the satellite's orbit to within several meters of accuracy. These coefficients are determined at the ground master control station by solving an inverse navigation problem. New coefficients for the ephemeris are uploaded to each satellite daily. The coefficients are stored with greater precision in the satellites than in the GPS receiver, so new values must be downloaded every hour to the receiver. The receiver precalculates interpolating polynomials with respect to the receiver time from the coefficients in the ephemeris. These polynomials are interpolated at specified times to determine the satellites' position.

User Segment

The User Segment consists of the unlimited population of passive receivers which use triangulation with four satellites to obtain a position fix. As each satellite continuously broadcasts its unique pseudo-random stream of binary 1's and 0's, the user's receiver produces an on-board replica of the bit stream, and, by means of spread spectrum correlation, determines the time delay between the time the signal is transmitted and received (2). By multiplying the time delay from each satellite by the speed of light, as shown in Figure 3, the range from satellite to user is determined ($r = c \times \Delta T$). This is called pseudo-range because it is not compensated for the user's clock error.

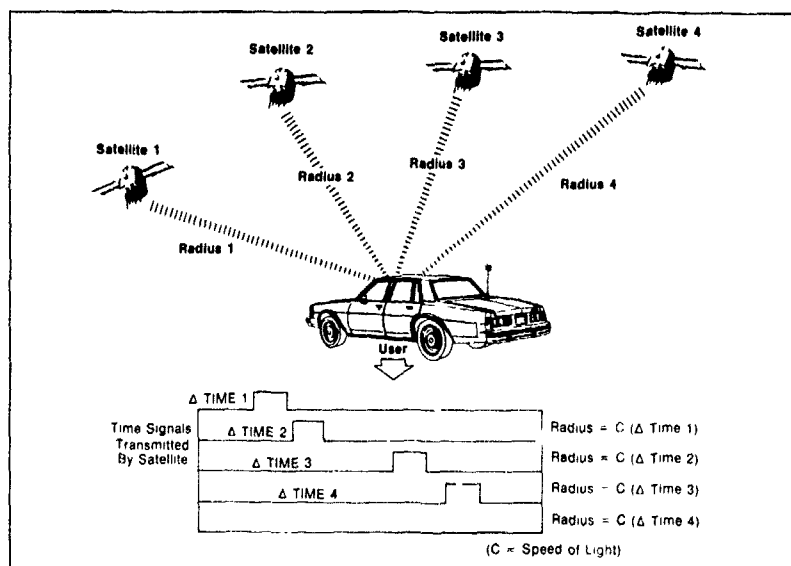


Figure 3 - NAVIGATING WITH THE GPS

If the receiver's clock were synchronized with that of the satellites, then three measurements of this type would pinpoint the user's location; but, because the user clock may be off by several milliseconds, a measurement is taken from a fourth satellite. This allows the receivers to use inexpensive crystal clocks instead of the precise atomic standards used in the satellites. Measurements to four satellites allows the computer to formulate four simultaneous equations which, as shown in Figure 4, eliminates the clock bias (C_b). Subsequent computation of user coordinates U_x , U_y , U_z are then converted to user latitude and longitude.

The software controlling the GPS receiver is motivated by a functional description of the navigation process. The purpose of navigation is to determine one's position within a given coordinate system. In this case, a cartesian coordinate system is chosen for the internal navigation calculations. This coordinate system is termed Earth Centered Earth Fixed (ECEF) because its Z axis is co-linear with the earth's polar axis and its X axis passes through the Greenwich Meridian. The navigation solution is transformed from ECEF coordinates to geodetic coordinates (latitude, longitude, and altitude) for the convenience of the user. This transformation is made using an ellipsoidal model for the earth. A variety of ellipsoids have been used in cartography, and the GPS system uses the coefficients for the WGS-72 (World Geodetic Survey) geocentric ellipsoid which was derived from satellite measurements.

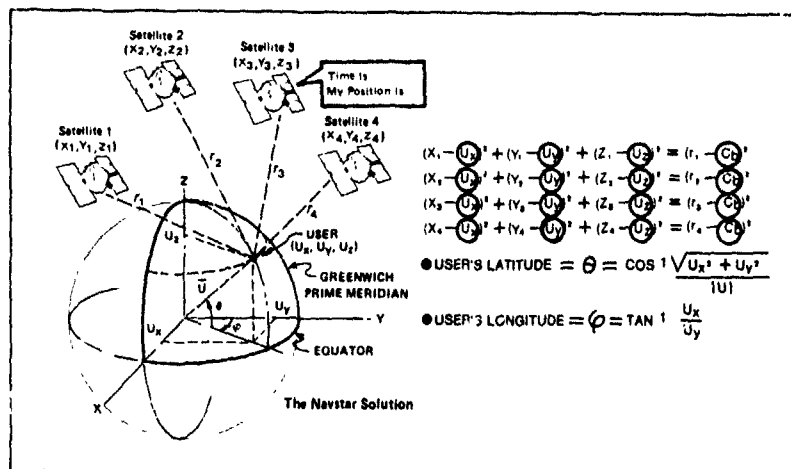


Figure 4 - GPS NAVIGATION EQUATIONS

Different applications dictate different receiver architectures. For example, tracking the dynamics of an F-16 fighter requires simultaneous processing of the signals from all four satellites resulting in four receiving channels, each locked to a separate satellite. A fifth channel tracking the next most appropriate satellite (based on signal/noise, elevation angle, and other considerations) is used to maintain continuity of the navigation solution should one of the four descend below the horizon.

However, to track a vehicle of relatively low dynamics, such as an automobile, requires only a single channel receiver which is time-shared to sequence among the four satellites.

The navigation signals are transmitted from each satellite at two L-band carrier frequencies: L_1 at 1575.42 MHz and L_2 at 1227.6 MHz. This permits corrections to be made for ionospheric propagation delays because the change in the velocity of a radio wave passing through the ionosphere is frequency dependent, making measurements at two frequencies enables the system's computer to eliminate this bias.

The signals are modulated with two pseudo-random noise codes: P, which provides precise measurement of time and position and is currently reserved for military applications, and C/A which provides for quick acquisition of the satellite signals. Because of the low dynamics associated with land vehicles and the modest tracking accuracy requirements, receivers such as GM's prototype are designed to receive only the L_1 carrier using the C/A code. Both L_1 and L_2 also are modulated with 50 bps data relevant to each satellite's status and orbital parameters.

The signal encoding accomplishes two basic and primary functions in system operation: it enables identification of each satellite because the code pattern is unique for each one; it also enables the receiver to determine the signal propagation delay by measuring the phase shift required for the receiver to match an onboard code of the same pattern.

The wideband nature of the signal (spread spectrum; C/A code 2.046 MHz bandwidth, P-Code 20.46 MHz) combined with the complex code modulation, the relatively low radiated satellite power levels, the extreme user-to-satellite distances, and the fact that both satellite and user are moving creating large Doppler excursions on the carrier frequencies, results in signals which are imbedded in ambient noise and very difficult to detect.

Detection is accomplished, however, using well known spread spectrum techniques which compress the large incoming signal bandwidth down to system band-widths of only a few hertz (3). This allows adequate signal-to-noise ratios to extract the signal. Tracking filters are used to keep the carrier and code in lock during the large Doppler shifts of the carrier.

The carrier tracking loop varies the reconstructed carrier frequency to match the incoming carrier, and the code tracking loop tracks the peak of the correlation function by controlling the shift of the local code replica.

In particular, all of the satellites broadcast on the same carrier frequency of 1575.42 MHz, but each satellite phase modulates this carrier with a different digital pseudo-random noise (PRN) code. This modulation spreads the carrier sufficiently to reduce the power level at the carrier frequency below the thermal noise level of the receiver. The pseudo-random codes are stored in the receiver, and they must be known in order to compress the signal bandwidth and recover the data modulation on the carrier. The receiver must also anticipate the doppler shift of the satellite signal in order to lock onto the carrier. This shift can vary by ± 6 KHz depending upon the satellites position in the orbit. The doppler shift must be known within 500 Hz in order for the carrier tracking loop to lock onto the signal, and this doppler shift is anticipated from the almanac by calculating the satellite's velocity component in the direction of the user. Given the doppler shift, the carrier lock circuitry is prepositioned to the expected signal frequency. The distance to the satellite is also anticipated from the almanac by differencing the satellite and estimated user positions. Since the propagation delay between the satellite and the user is equivalent to a shift of the PRN code, then the code shift may be anticipated from the estimated satellite distance in order to aide the code lock loop. Both the carrier frequency and the code shift are tracked by hardware in the front-end of the receiver. The carrier is first prepositioned, and then the code tracking loop is dithered until the known PRN code best correlates with the incoming signal. The correlation is found from the integral of the product of the signal with the PRN code over the duration of the code. On a cold start, the nominal carrier frequency is also dithered until either the signal is captured, or all combinations of carrier/code shift are exhausted.

SYSTEM FUNCTIONAL DESCRIPTION

The functional block diagram of the current navigation system is illustrated in Figure 5. The navigation system consists of 1) navigation sensors, 2) user interface, and 3) the display computer. Three different navigation sensors have been included. The GPS and LORAN receivers are self contained units, mounted in the trunk of the car, while the dead reckoning sensor is derived from combined measurements of a flux-gate compass, mounted below the roof of the car, and the odometer. The user interface consists of a series of switch inputs, the map cartridge reader, and the CRT display. Push-button switches and a joystick are mounted in the central armrest in the front seat. A thumbwheel switch, an antennae selector switch, and the CRT display have been mounted in the dash. A map cartridge reader is mounted below the dash. The display computer, consisting of an interface board, a CPU board, and a graphics controller board have been mounted in a card-cage located in the trunk of the car.

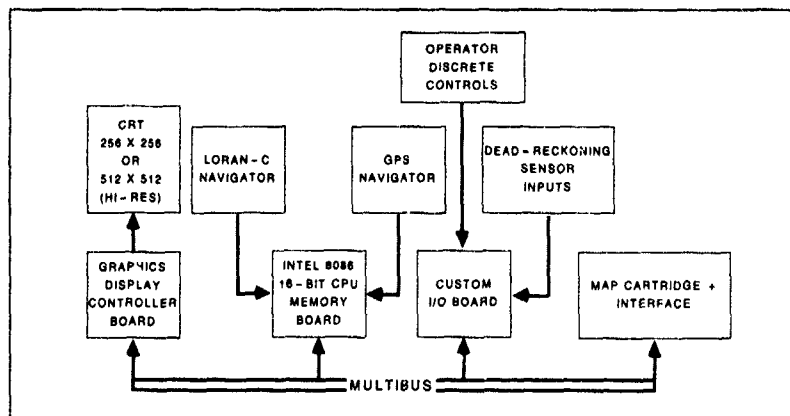


Figure 5 - CURRENT NAVIGATION SYSTEM FUNCTIONAL BLOCK DIAGRAM

For developing a system adequate for land vehicles and to permit software flexibility in the design of various acquisition and tracking algorithms, a GPS receiver configured in an algorithm development system environment was packaged. The development GPS receiver/processor architecture is shown in Figure 6.

The GPS receiver communicates with both the map display computer and with an operator Control/Display Unit (CDU). Normally the GPS receiver will commence a warm start sequence when the vehicle is powered up, and knowing its last position and the current time, the receiver will consult the almanac to select which satellite would be most visible, and then attempt to acquire that satellite. If the satellite antenna is unobstructed, and the receiver and satellites are working properly, then the satellite will be acquired, after which the remainder of the satellites in the constellation will be acquired, and the receiver will commence navigation, sending calculated GPS position to the display computer once per second. The user's current position may also be downloaded from the display computer to the GPS receiver by pressing the ENTER key on the operator console after positioning the display cursor on the map at the current vehicle position.

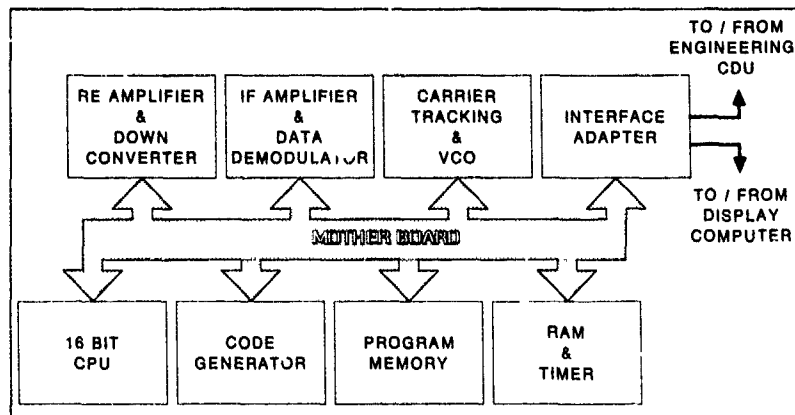


Figure 6 - GPS DEVELOPMENT SYSTEM FUNCTIONAL BLOCK DIAGRAM

GM's current test and demonstration vehicle is equipped with three navigation systems, as shown in Figure 7.

- NAVSTAR Global Positioning System
- LORAN-C
- FLUX-GATE compass and odometer for dead reckoning

This test configuration permits the relative evaluation of three independent or integrated approaches. The seven satellites currently deployed limit the navigation window to approximately 6-7 hours per day.

The Navstar solution data stream is linked to the display computer, (a 16 bit microprocessor) for decoding and correlating with digitized roadmap data which has been referenced to latitude and longitude coordinate frame of reference. In this manner, the vehicle location is pinpointed against a geographic map directory in memory and used to drive the color CRT with vehicle location and surrounding streets, roads and highways.

The display also has an operator-selectable "trip planning" mode which allows the driver to pre-plan a trip by inserting origin, destination, and intermediate stopping points, all of which are displayed at the touch of a button. Relative bearing and distance to each stopping point are then displayed while enroute. Other driver controls include multiple map scales to enlarge to increasing levels of detail, the streets and highways of a given area. The largest scale is automatically displayed when a destination point is approached within a 1-1/2 mile radius.

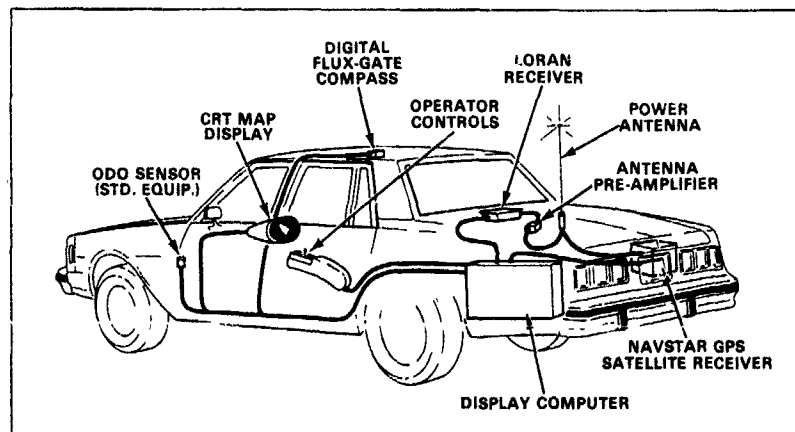


Figure 7 - CURRENT NAVIGATION SYSTEM FUNCTIONAL COMPONENT LAYOUT

The roadmaps are digitally stored in cartridges inserted in a slot in the car's instrument panel. The cartridges are changed as the vehicle travels from one remote area to another. Since a cartridge contains many map "pages", a new map is automatically drawn on the CRT as the vehicle, represented by the flashing rectangle, approaches the edge of the screen (automatic paging). The cartridges are programmed by digitizing U.S. Geological Survey maps.

Vehicle Testing

Development testing for the purpose of GPS software debugging and tracking algorithm refinement was accomplished by implementing the test van shown in Figure 8. The van included data monitoring and storage equipment in addition to the vehicle location sensing and display equipment. The equipment was tested on the roads at Pike's Peak to evaluate the system's sensitivity to rapidly changing altitudes and the effect of signal blockages in mountainous terrain. Under these conditions, the average position errors were within the expected range of less than 50 meters. The RMS position error for all tests combined were less than 100 meters.

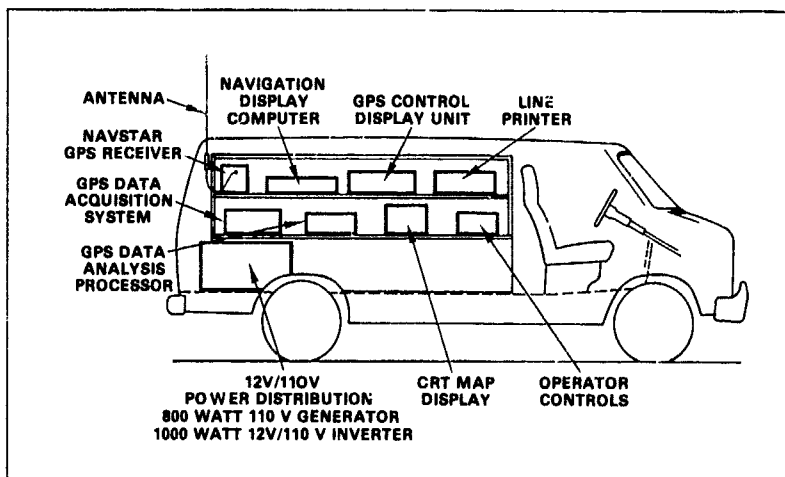


Figure 8 - NAVIGATION TEST VAN FOR GLOBAL POSITIONING SYSTEM

LIST OF REFERENCES

1. D. W. Henderson and J. A. St-ada, "NAVSTAR Field Test Results" Global Positioning System Collection of Papers Published in NAVIGATION and Reprinted by the Institute of Navigation Vol. ION 0-936406-003, pp. 234-246.
2. Thomas A. Stansell, Jr., Director, Advanced Programs, "Meeting the GPS Challenge" Reprinted from Magnavox's February and May 1983 issues of POINTS and POSITIONS.
3. N. B. Hemesath, "Performance Enhancements of GPS User Equipment", Navigation: Special Issue on GPS/Journal of the Institute of Navigation, Vol. 25, No. 2, Summer 1978.

ACKNOWLEDGEMENTS

The author wishes to acknowledge the efforts of Rockwell-Collins Avionics Group under the guidance of Dr. N. B. Hemesath for development of the GPS receiver hardware and software; and to Hunter Systems Company, Inc. under the direction of Tom Hunter for development of the display, display processor and associated software. Both organizations performed these tasks while under contract to GM.

AN INTEGRATED SYSTEM FOR LAND NAVIGATION

J.C. McMILLAN

DEFENCE RESEARCH ESTABLISHMENT OTTAWA
OTTAWA, ONTARIO
CANADA K1A 0Z4**SUMMARY**

Conditions for land navigation are among the most severe in the arctic, where until GPS becomes fully operational, there will be no single system capable of continuously providing the necessary position and heading accuracy. Even when GPS is available reliability considerations will dictate that a self contained or autonomous backup be available, certainly for the military user. The Defence Research Establishment Ottawa (DREO) has therefore developed a multi-sensor optimally integrated navigation system to satisfy the present operational requirements of the Canadian land forces. This system, called PLANS (Primary Land Arctic Navigation System) was designed and built to be a highly reliable, moderately accurate and moderately priced, nonradiating, automatic navigation system for all weather off the road use. Although designed primarily to meet an arctic requirement, PLANS would of course be just as applicable for desert navigation, or for any application in which reliability and accuracy are a priority, or where simpler and less costly methods are ineffective.

PLANS combines several self contained sensors with two satellite receivers, using an 8 state Kalman filter on a 68000 based microcomputer, to provide a continuous optimal estimate of position, height and heading. The self contained sensors consist of a strapdown gyrocompass/directional gyro unit, an odometer pickoff for speed, a magnetic fluxgate sensor (with a detailed geomagnetic field model in the software) and a baroaltimeter (with a digital terrain elevation map providing assistance). The satellite receivers are a single channel C/A code GPS and dual channel Transit. This section will describe the requirement, the difficulties in meeting this requirement, the design of a multi-sensor integrated system solution, some simulation results and some field trial results.

1. PROBLEM STATEMENT

Perhaps the most stringent requirement for land navigation is imposed by winter operations in the high arctic. The serious consequences of being lost or failing to find a supply dump in the arctic make reliability and accuracy vital necessities. Because of weather related poor visibility conditions, it may also be necessary to navigate by instrument very close to the destination in order to actually find it. Furthermore, a combination of circumstances conspire to make the arctic particularly difficult area in which to navigate by conventional means.

The Canadian arctic, from 60° to 85° latitude, consists of about 7,000,000 square kilometres, roughly half of which is land. To the untrained eye of a non-inhabitant most of this area appears completely barren and featureless, especially during the long winter period where land-water division is obscured. Certainly above the tree line any movement beyond line of sight requires some non-trivial navigation capability. There are generally no rail or road systems as we know it, and in most areas no permanent recognisable landmarks. The normal map reading skills are therefore completely inadequate. Frequent extended periods of low visibility also interferes with visual navigation, and make celestial navigation unreliable at best. In any case sextants and sun compasses are not sufficiently accurate.

In many ways arctic navigation is similar to desert navigation, except that there are several added complications. The area in question is unique in containing the north geomagnetic pole, which makes the use of a hand-held magnetic compass futile over most of this area. This is because the geomagnetic field lines at the magnetic pole are vertical, and the compass relies on the horizontal component of the field. As shown in reference [1] random fluctuations in the direction of the field vector about its mean value will induce heading errors of magnitude inversely proportional to the horizontal field strength. Figure 1 shows how this magnetic heading error magnitude varies with latitude along two selected longitude lines, -100° (which virtually passes over the magnetic pole), and -60°. This figure represents the error of a stationary magnetic compass, assuming a constant level of field fluctuation, and that the local mean field is known, that the disturbing field from the vehicle is known, and that the accuracy at the equator is 1°. In fact the level of fluctuations in the geomagnetic field are not constant, but show a definite dependence on geomagnetic latitude and time of day, as shown in Figure 2 (borrowed from [2]).

Of course the usual radio navigation aids do not extend coverage through the arctic, with the possible exception of Omega [3], which is not sufficiently accurate. The Transit satellite system does provide some position fixing capability, if velocity and height can be continuously provided. Since the Transit satellites are in polar orbits, fixes can be obtained more frequently at higher latitudes, with more than one fix per hour expected above 60°. The accuracy of a Transit position fix depends upon the accuracy of the velocity provided, but can be quite adequate, especially with a 2 channel receiver (to remove the ionospheric effect) in stationary mode. GPS position and velocity is quite adequate when it is available, even with a C/A code receiver. Until it is fully operational however, the gaps in coverage prevent total reliance on GPS.

This therefore still leaves the requirement to dead reckon (DR) between Transit fixes or through GPS gaps, so that the velocity determination cannot be avoided. Now the length of the velocity vector, or the speed, is easily estimated by using the vehicle odometer information, but the heading presents some difficulty. The first alternative to magnetic direction finding is gyrocompassing which uses the fact that the horizontal component of the earth's rotation rate ω is in the north direction. Gyrocompassing becomes more difficult at higher latitudes λ , where the horizontal earth rate ($\omega \cos \lambda$) decreases quickly as one moves north. Thus the gyrocompassing accuracy of a given sensor varies as the secant of the latitude. In practice of course for each gyro there is a latitude above which it will not settle at all. For the

moderately priced type of unit suitable for land vehicles this is typically given as about 80°. Fortunately, the present location of the geomagnetic pole is such that the land area where magnetic heading is not available (using a magnetometer, not a hand held compass) probably does not intersect the land area where gyrocompassing is not possible, as shown in Figure 3. Of course neither of these regions are sharply delineated in fact. Rather each heading measurement become increasingly less accurate moving towards and into its corresponding region. The point at which each measurement becomes useless depends on the particular sensors chosen, how they are used, and the magnetic conditions at the time. The important point to note is that in the extreme northwest, which is the area of greatest difficulty, the two methods complement each other, with magnetic heading improving as gyrocompass heading deteriorates and vice versa.

Fig. 1. Magnetic Heading Error Sensitivity

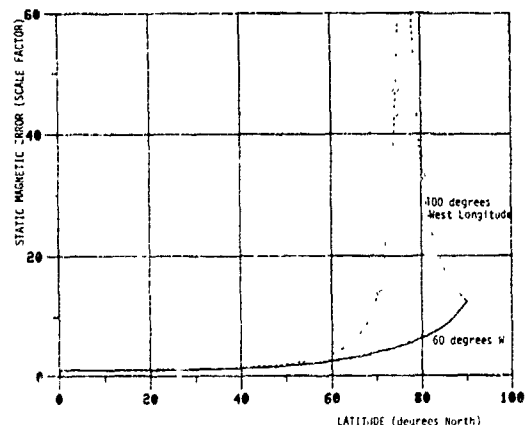


Fig. 2. Mean Hourly Magnetic Field Fluctuations

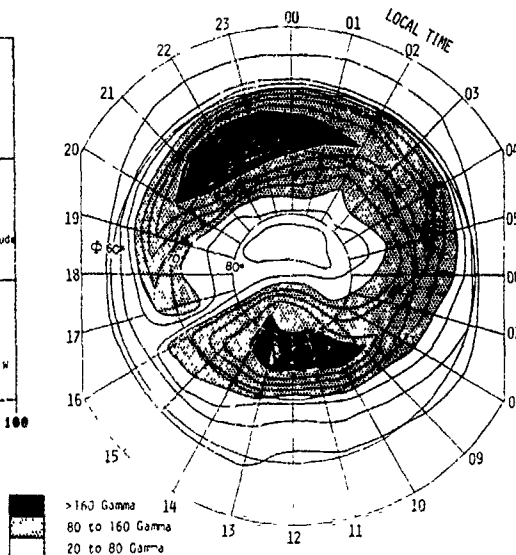
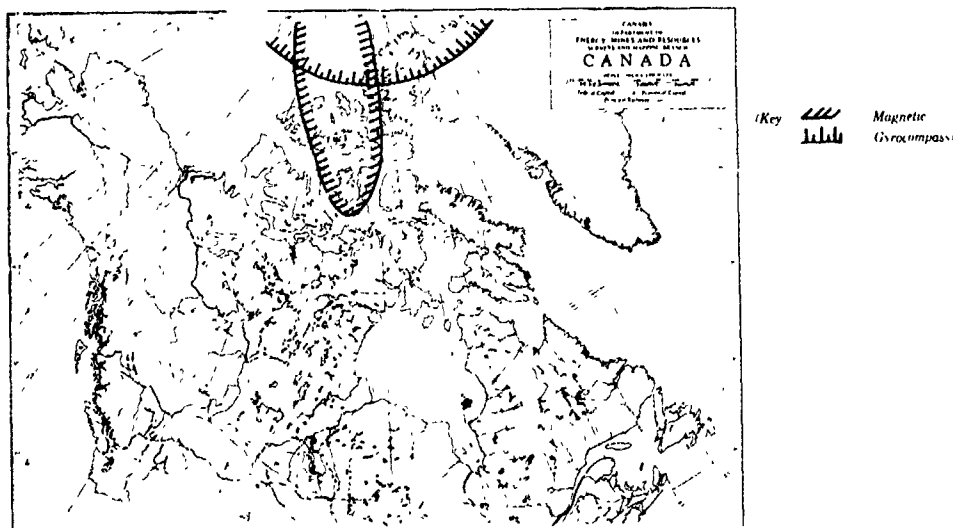


Fig. 3. Region of Magnetic and Gyrocompass Heading Difficulty

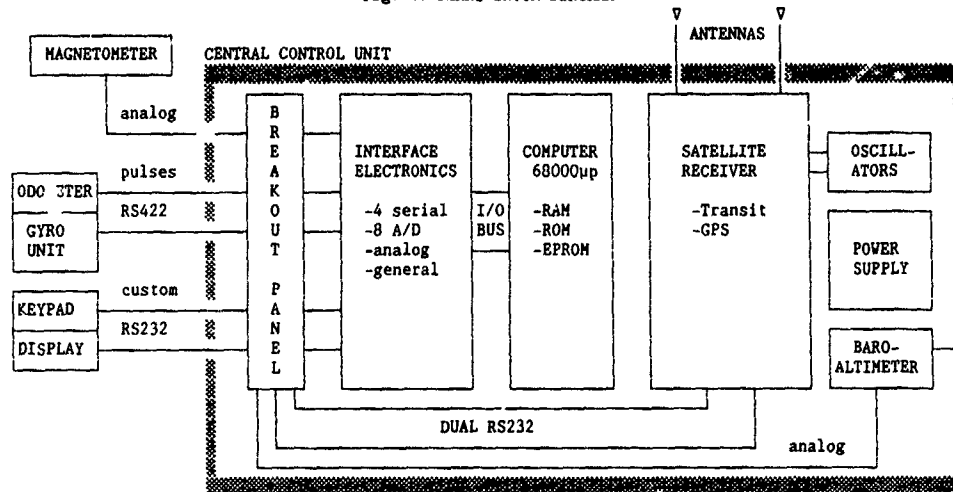


2. SYSTEM SOLUTION

With the problem as stated above, the obvious solution is to combine one or both of the satellite receivers with a magnetic sensor, a gyrocompass/directional gyro unit and an odometer pickoff, in an optimally integrated system, using a Kalman filter to properly exploit the strengths of each sensor. After investigating some non-conventional alternatives, such as VLF direction finding and short baseline interferometry using Transit to obtain heading information [4] DREO decided that an integrated system with off-the-shelf components was the preferred solution. Since no such system was available, it was necessary to develop one. Previous experience in the design and development of a Kalman filter based integrated navigation system [5],[6] was influential in this decision, and in the design approach. The major components chosen for PLANS are shown in Figure 4, in block diagram form.

This design is actually a transitional configuration, since PLANS was first built without the GPS receiver, and when GPS is fully operational the Transit receiver can be eliminated. Since the Transit receiver requires height input, a baroaltimeter has been included, which is augmented with a small terrain elevation database. These are appropriately blended in the Kalman filter. Use of a Transit height measurement to calibrate the barometric height was considered, but would have required the vehicle to remain stationary through several satellite passes plus some additional receiver capability, with its attendant cost. The VRU heading can be combined with speed obtained from the vehicle odometer to form the velocity vector. A deadreckoning solution can then be obtained by integrating this velocity. The core of the integrated system is this deadreckoning system, consisting of the VRU and odometer. All other sensors can be thought of as aiding sensors.

Fig. 4: PLANS BLOCK DIAGRAM



The magnetometer provides additional heading information, which is combined with the VRU heading through the prefilter and the Kalman filter, as described below. The magnetic measurement of course is compensated to account for the difference between true north and the local magnetic field direction. This requires a geomagnetic field model and a vehicle field model. The geomagnetic model used in PLANS is the IGRF85 model (International Geomagnetic Reference Field, 1985) [7]. The vehicle field model is a second order Fourier series representing the vehicles permanent field ("hard iron") and induced field ("soft iron") and is found by using the VRU to calibrate the installed magnetometer while driving the vehicle over the complete range of headings [1].

The gyro unit and magnetometer are not the only sources of heading information available. Of course GPS velocity can be used when the vehicle is moving and GPS is available, but it is also possible to obtain heading information from a Transit position fix, in an indirect way. This is due to the known sensitivity of the Transit fix to the error of the velocity fed to the receiver. The Kalman filter, using a detailed model of this relationship, can then update its estimate of the velocity error (including heading error) when it processes the Transit position fix. In simulation this technique is quite successful, however, it does rely on the Transit receiver providing correct satellite elevation angle and direction of travel information.

3. SENSOR SELECTION

The gyro unit used in the proof of concept study was a strapdown "Vehicle Reference Unit" (VRU) having one tuned rotor gyro (2 axes), one single axis rate gyro and 2 accelerometers. This unit can gyrocompass in about 5 minutes (longer at higher latitudes) if the vehicle is stationary and pointing within about 30° of north. This provides an initial heading, accurate to about one degree secant latitude (1σ), after which the VRU automatically switches to directional gyro mode and the vehicle can be moved.

For the PLANS proof of concept development model a two axis magnetic compass sensor unit was used which incorporates a gimbaled fluxgate sensor assembly capable of directly measuring the horizontal components of the earth's magnetic field. The manufacturer claims that it will provide magnetic heading accurate to 2° for magnetic dip angle from 0° to 80° which includes a large portion of northern Canada. Because of this and a potential low temperature problem, a more suitable 3-axis strapdown magnetic sensor unit has been obtained for the next phase of development.

The baroaltimeter sensor chosen was a High Output Pressure Transducer with gage pressure range rated at 0 to 25 pounds per square inch (0 to 172 kilo-Pascals). The range of interest is roughly 11 to 16 psi corresponding to a barometric height range from about -700 metres to 2,500 metres. This barometric sensor weighs only about 100 grams and occupies only about 100 cc.

The satellite receiver chosen for the PLANS development model was a commercial 2-channel Transit receiver (intended for the marine environment) which provides convenient serial output and the ability to be easily upgraded to include a C/A-code GPS receiver. The keypad, display and case were discarded, and the

necessary electronics (including battery backup) was repackaged with a 68000 based computer system, the interface cards, power supplies and the baroaltimeter. In this way the satellite receiver(s) can be completely controlled by the PLANS processor, using a keypad emulator.

4. ERROR MODELLING

In order to properly integrate these different sensors using a Kalman filter, all significant stochastic and deterministic errors must be carefully modelled. Deterministic errors, such as magnetic declination, are modelled so that their effects can be directly removed as much as possible, and the remaining errors are modelled as random processes to enable the Kalman filter to estimate them and to calculate the appropriate gain matrix to weight the various measurements. An overview of the modelling process for PLANS is given here. It should be noted that to a large extent this error modelling is generic, with the form of the models depending only on the type of sensors being used. The actual sensor models chosen will of course determine the numerical values of the parameters, enabling the PLANS software to be easily adapted to different qualities of sensors. This flexibility allows the cost/performance tradeoff and final sensor selection to be made late in the development program, a necessary measure since PLANS system development began before sensors with the desired qualities were available.

The types of stochastic models used in the PLANS Kalman filter are limited for practical reasons to first order Markov processes and uncorrelated white noise [12].

4.1 DEAD RECKONING ERRORS

Dead reckoning (DR) is the determination of the position at time t_1 by knowing the initial position at time t_0 and integrating the velocity vector from t_0 to t_1 . For land and sea navigation the velocity vector is generally obtained by measuring the heading θ , speed S and if possible pitch ϕ . For PLANS the DR heading and pitch is measured using the VRU and the speed is measured using the vehicle odometer.

To obtain this speed measurement a speed sensor was installed on the odometer cable which generates the pulses per revolution used to measure the along-track distance travelled by the vehicle. These pulses are fed directly to the PLANS computer which counts them over the integrating interval Δt . This pulse count C is then multiplied by the appropriate scale factor F , to obtain the distance D , moved over the ground during the time interval Δt . The vehicles average speed over this interval is therefore $S = D/\Delta t = CF/\Delta t$.

The VRU heading measurement is assumed to be the clockwise angle θ from true north to the horizontal velocity vector. If the velocity vector is of length S , and is pitched out of the horizontal plane by the angle ϕ , then the north and east dead reckoning velocity components are:

$$\begin{aligned} V_n &= \frac{CF \cos \phi \cos \theta}{\Delta t} \\ V_e &= \frac{CF \cos \phi \sin \theta}{\Delta t} \end{aligned} \quad (1)$$

The nominal value of the scale factor F is provided by the manufacturer, but to achieve the necessary accuracy for PLANS, a more precise value was determined by means of a simple calibration run. The dead reckoned position is then obtained by integrating these velocity components over the WGS 84 ellipsoid, as follows:

$$\lambda_{t+\Delta t} = \lambda_t + \frac{V_n \Delta t}{R_n + h} \quad (2)$$

$$L_{t+\Delta t} = L_t + \frac{V_e \Delta t}{(R_e + h) \cos \lambda}$$

where h is the height of the vehicle above the ellipsoid, and R_n and R_e are the meridional and prime radii of curvature of the ellipsoid:

$$R_n = \frac{A^2}{B(1 + E \cos^2 \lambda)^{3/2}} \quad (3)$$

$$R_e = \frac{A^2}{B(1 + E \cos^2 \lambda)^{1/2}}$$

where

$$E = \frac{A^2}{B^2} - 1 \quad (4)$$

and A and B are the semi-major and semi-minor axes of the ellipsoid. There are many possible sources of error in this DR calculation most of which can be categorised as speed or track errors. The important errors are expected to be as follows:

"Speed" errors: s1 - vehicle track slippage (along track)

- s2 - odometer scale factor error
 - s3 - computer clock error
 - s4 - measured pitch error $\delta\theta$
 - s5 - odometer pickoff sensor fault
 - s6 - data communication fault
 - s7 - invalidation of odometer when vehicle is used in amphibious mode, or on drifting ice.
- (5)

- Track errors:
- t1 - gyro misalignment (not exactly along vehicle axis)
 - t2 - gyrocompassing error
 - t3 - directional gyro error
 - t4 - vehicle lateral track slippage (crabbing)
 - t5 - discretisation error (resolution)
 - t6 - data communication fault
 - t7 - invalidation of heading as track when in amphibious mode, or on drifting ice.
- (6)

The most significant of these DR error sources are expected to be s1, s4, t1, t2 and t3. Although other errors such as s7 and t7 are potentially more serious, they are much less likely to occur.

The vehicle track slippage s1, during acceleration, deceleration and moving on a grade, is expected to produce approximately 1% error in distance travelled, which is converted in PLANS to speed. Treated as a stochastic process, this error is continuous, bounded and nominally zero mean. It therefore can be modelled to first order (which is quite adequate for this purpose) as a first order Markov process (FOMP).

Careful calibration on the host vehicle, to adjust the variable scale factor, can reduce the odometer scale factor error s2 to a comparably negligible level.

If the vehicle pitch angle θ is not available, it would have to be approximated by zero in equation (1). In this case the effect of pitch error s4, incurred by neglecting the non-horizontal component of the velocity, is a scale factor equal to $(1 - \cos\theta)$. This is an error of less than 1% of distance travelled for pitch angles of less than 8° , which is a grade of about 1 to 7. At steeper grades this error increases fairly rapidly (about 4% error at 16° pitch for example). Stochastically this error is random, bounded, continuous and zero mean, corresponding to a FOMP just as s1.

The heading error consists mainly of the static error (t1+t2) plus the dynamic error, either t2 (if a dynamic gyrocompass is used) or t3 (if a directional gyro is used). The static error is largely due to the difficulty of accurately aligning the gyro sensitive axis to the vehicle's forward movement axis during installation. In practice the gyro is mounted by aligning a mark on the gyro housing, representing the sensitive axis, to a mark on a mounting plate, representing the vehicle's forward axis. The difficulty in practice is accurately installing the mounting plate in each vehicle. By using a software heading offset, careful calibration on the host vehicle can reduce the effective misalignment to well below one degree.

The initial gyrocompass heading error t2 is expected to be below 1° (1σ) at low latitudes, but increase with latitude as $\sec\lambda$. This error will be zero mean, continuous and bounded: again a FOMP (but not quite stationary since λ will slowly change).

In directional gyro mode the heading error t3 will vary slowly in an unbounded manner. The drift rate however, will be bounded, continuous and nominally zero mean, so that the heading error rate can be modelled as a stationary FOMP. This drift rate will be unaffected by latitude.

4.2 MAGNETIC HEADING ERRORS

The magnetic flux valve measures the direction of maximum horizontal magnetic field strength. Since the earth's magnetic field is roughly aligned with its axis of rotation, on a global scale the north magnetic pole is close to the geographic north pole. Therefore at low latitudes the magnetic direction is approximately north. Unfortunately however, the north magnetic pole is almost centrally located in the Canadian arctic, so that the magnetic field direction differs substantially from true north throughout this area. The difference between true north and magnetic north is known as the Magnetic Variation, or the Magnetic Declination.

To obtain a usable magnetic heading it is therefore necessary to apply the magnetic declination correction to the measurement. The geomagnetic field model selected is known as IGRF85 (International Geomagnetic Reference Field 1985) [7], which is a global spherical harmonic model of degree and order 10 in the main field and 10 in the secular field. A Fortran program based on this model is available (from the American National Oceanographic and Atmospheric Administration for example), which computes the horizontal field strength, the dip angle, the total field strength and its three Cartesian components (from which the declination can be calculated) as well as the rate of change of each of these quantities. The dip angle and horizontal field strength are useful in predicting the accuracy of the magnetic flux valve measurement.

The error in magnetic heading is largely due to unpredictable temporal variations in the direction of the local magnetic field vector, but there can also be significant sensor errors, especially if a pendulous gimbal is used to keep the sensor horizontal. The most important factors affecting magnetic heading accuracy are as follows:

- m1 - local variations in the geomagnetic field (temporal and spatial)
 - m2 - magnetic fields induced in the vehicle by the earth's field
 - m3 - permanent fields in the vehicle
 - m4 - sensor misalignment error
 - m5 - dynamically induced deviations from the horizontal (gimbal systems)
 - m6 - low horizontal field strength
- (7)

The uncompensated magnetic heading error can then be expressed as:

$$\theta - \theta_m = m_1(\lambda, L, t) + m_2(\theta) + m_3(\theta_m) + m_4 + m_5(t) \quad (8)$$

where θ is the geographic heading and θ_m is the measured magnetic heading.

The most obvious source of error in magnetic heading is due to the fact that the magnetic field vector does not generally point towards the geographic north. As described above, the large scale spatial component of m_1 can be modelled fairly well. On a local scale the magnetic field is affected by permanent or induced magnetic fields in the vehicle itself or some nearby structure (manmade or geological). The permanent field of the vehicle m_3 , will add vectorially to the earth's magnetic field vector, introducing a heading error that varies sinusoidally as a function of the vehicle's geographic heading, with a period of 360° . This is often called the "hard-iron effect", and is due to magnetised portions of the vehicle or its load. If the load effect can be neglected then this error can be compensated for by knowing its amplitude and phase. The induced field m_2 , known as the "soft-iron effect", is due to the high permeability portions of the vehicle (such as iron and steel) warping the earth's magnetic field. This error is similar to m_3 in that it can be compensated for by determining its amplitude and phase, but different in that it is a function of magnetic heading and that m_2 has a period of 180° . If we let $G(\lambda, L, t)$ be the magnetic declination predicted by the geomagnetic field model, (a function of latitude longitude and time) then equation (8) can be written as

$$\theta - \theta_m = G(\lambda, L, t) + \Delta m_1(t) + a_2 \sin(2\theta_m + \delta_2) + a_3 \sin(\theta + \delta_3) + m_4 + m_5(t) \quad (9)$$

where $\Delta m_1(t)$ is the error of the field model, a_2 , a_3 , δ_2 and δ_3 are constants associated with the vehicle's permanent and induced field and m_4 is the sensor misalignment. These constants can all be found by a fairly simple calibration procedure. After the geomagnetic and vehicle field models have been applied, the remaining magnetic heading error is

$$\theta - \theta_m = \Delta m_1(t) + m_5(t) \quad (10)$$

where field uncertainty $\Delta m_1(t)$ will be randomly varying in a continuous manner, much like a first order Markov process, and the dynamically induced error m_5 will be largely uncorrelated, and can be treated as white noise.

Although a good geomagnetic field model can correct for the spatially varying component of m_1 , there will remain a significant time varying component, especially in the vicinity of the magnetic poles. This temporal fluctuation in the field is the most serious problem in obtaining an accurate magnetic heading, especially if it is magnified by the geometric effect near the poles where the field has a large vertical component and a small horizontal component. This near vertical alignment makes it very difficult, if not impossible, for a flux valve to obtain any heading measurement at all. This is primarily because the change in the magnetic declination θ , and hence in magnetic heading θ_m , as a function of temporal changes in the north and east components of the magnetic field δX and δY , is inversely proportional to the horizontal field strength H :

$$\delta\theta = \frac{-\sin\theta \delta X + \cos\theta \delta Y}{H} \quad (11)$$

where $\delta\theta$ is the change in the magnetic declination, in radians. Equation (11) is a purely geometric relationship, found by differentiating the definition $\theta = \arctan(X/Y)$.

These horizontal magnetic field components X and Y can exhibit substantial temporal variation, especially in the far north. Although these variations are random, their expected magnitude does vary in a fairly regular fashion with time of day and with season. There is also a correlation with the 11-year solar cycle and the 27-day period of the sun's rotation. The mean hourly variation of these field components was shown in Figure 2 for an area around the north magnetic pole (the coordinates are Geomagnetic latitude, which is centred on the magnetic pole). Over a large portion of the Canadian arctic, this average variation is typically on the order of 100 Gammas, but the actual variation can at times easily be as much as 500 γ . The horizontal field strength H is generally less than 17,000 γ in the Canadian Arctic, resulting in declination changes of more than 3° , and in some areas more than 10° (see [2] for more details).

Another source of magnetic heading error m_5 is also strongly effected by the horizontal field strength. This is due in part to the problem of keeping the flux valve vertically aligned. Any small misalignment will cause the strong vertical component of the field to project into the flux valve's sensitive (nominally horizontal) plane and introduce a large unpredictable error. This error is described in [8] as a function of an acceleration that is assumed to have caused the pendulous gimbal to tilt. For small accelerations the resulting heading error is approximated in [8] by an expression equivalent to:

$$m_5 = \frac{A Z \sin \beta}{g H} \quad (12)$$

where Z and H are the vertical and horizontal components of the geomagnetic field, A is the small acceleration, g is the gravitational acceleration and β is the angle between the acceleration vector and

north. This is clearly sensitive to small horizontal field strength.

These dynamically induced errors were examined in more detail as a function of gimbal pitch and roll angles θ and ψ , and magnetic field components X, Y, and Z (north, east and down). The exact expression is rather long, but for a small pitch angle θ and zero roll, the heading error is approximately

$$m5 = \tan^{-1} \left[\frac{\theta Z (X \sin \theta + Y \cos \theta)}{H^2 - \theta Z (X \cos \theta - Y \sin \theta)} \right] \quad (13)$$

As would be expected, the argument of this expression has singularities at pitch angles which place the magnetic field vector perpendicular to the gimbal plane (the sensitive plane of the sensor) which is nominally horizontal. In the Ottawa area for example, this singularity occurs at a pitch angle θ of only 18.4° when the heading θ is zero. In other words a gimbal pitch angle of only 18.4° or more would completely destroy the magnetic heading measurement accuracy.

It has been our experience with our gimballed magnetic flux valve that these dynamically induced heading errors are indeed quite serious during normal operation of the host vehicle, even over rather smooth roads. At a one Hz sampling rate this produced a heading error of several degrees that appeared largely uncorrelated in time. Of course the effect of this uncorrelated noise can be significantly reduced by implementing a simple prefilter.

4.3 VERTICAL ERRORS

The vertical DR position can be defined as the barometric altitude. The pressure-height relationship in the low altitude region (0 to 11 km), can be easily derived from the standard atmosphere equations described in [9] as follows:

$$P = P_0 (T/T_0)^{-g/aR} \quad (14)$$

$$T = T_0 + aY \quad (15)$$

where P is the measured air pressure, P_0 is the nominal sea level air pressure in the same units, T is the temperature, T_0 is the defined sea level temperature of the standard atmosphere (15.16°C), a is the standard atmosphere temperature gradient for altitudes less than 11,000 m (about -0.0065°C/m), g is the gravitational constant (9.80665 m/s^2), R is the gas constant ($288 \text{ joules/(kg}^\circ\text{C)}$) and Y is the vehicle height above sea level, in metres. Substituting (15) into (14) gives:

$$P = P_0 (1 + aY/T_0)^{-g/aR} \quad (16)$$

or

$$P = P_0 (1 + AY)^B \quad (17)$$

where A and B are known constants with approximate values:

$$A = a/T_0 = -2.2569 \times 10^{-5} \text{ meters}^{-1} \quad (18)$$

$$B = -g/(aR) = 5.2386$$

To obtain the barometric height Y from the measured pressure P, equation (17) must be inverted to yield:

$$Y = \frac{e^{\frac{\ln(P/P_0)}{B}} - 1}{A} \quad (19)$$

This barometric altitude is augmented by using a digital elevation map of the Canadian arctic, created at DREO using a memory-efficient symbolic storage technique. To create this map, height above sea level was manually read directly from the $1:1,000,000$ scale topographic maps. In this way an "average" height was visually estimated for the corners of each grid square, which are one degree in longitude by one half degree in latitude. This digital map covers the region north of 60° latitude and between longitudes 60° west and 140° west. Two dimensional linear interpolation is then used to calculate the correct height estimate for any location between these data points. This produces a height function that is continuous and that corresponds to the stored values on the grid corners.

The important vertical errors in the PLANS system are:

- v1 - inaccuracy of the pressure-height equation, due to weather
 - v2 - pressure transducer error
 - v3 - data communication fault
 - v4 - inaccuracy of the digital elevation map
- (20)

The vertical error v_1 requires some explanation. The pressure-height relationship given by equation (19) describes an idealised "standard atmosphere", whereas in reality changes in weather are constantly changing the air pressure. In the Canadian Arctic, normal pressure fluctuations are .1 to .2 kilopascals over a 12 hour period, and 2.5 to 5 kPa over a half week period. This would result in a false barometric height fluctuation of 8 to 10 metres over 12 hours and 200 to 400 metres over a half week. It is possible however for extreme weather to create pressure changes of 1 kPa per hour and 10 kPa per day or, equivalently, 80 metres per hour and 800 metres per day. This error is continuous, bounded and zero mean, and is modelled as a FOMP.

The pressure transducer error v_2 is claimed to be less than $\pm .11\%$ FS (full scale), and if it is calibrated, less than $\pm .02\%$ full scale plus $\pm .03\%$ FS/°C. This amounts to an uncalibrated error of less than 15 meters in barometric height, which is very small compared with v_1 , so the calibration is deemed unnecessary.

V_3 is meant to include any A/D conversion error, round off error, and any computational errors in evaluating equation (19).

Since the host vehicle is presumably restricted to the earth's surface, its height above the ellipsoid is in principle a well defined function of latitude and longitude. If a sufficiently detailed elevation map could be stored in the computer's memory, then the barometer would be unnecessary and the height error V_1 would not be very significant. However the map currently being used is one created at DREO with limited resources. In order for the height information to be coded and stored in an efficient manner it was necessary to quantize the heights. This vertical quantization is 50 metres for heights up to 1km, and 100 metres for greater heights. When the map is read as a function of position the heights from the surrounding four grid squares are linearly interpolated, producing a continuous function of position. The largest error in reading this map is due to the local deviation of the true height from the "area averaged" height, as visually estimated from the topographical charts. In other words, the limiting factor is the spatial resolution used to read the charts (about 60 km by 60 km) and the resolution of the charts themselves. We estimate that our map height error is on the order of 30 metres plus 10% of map height. This error is continuous, zero mean and bounded, and again modelled as a FOMP.

4.4 TRANSIT ERRORS

Since the DR position error will generally increase without bound, an independent position fixing system is required to periodically reset the DR position. Because of its accuracy and coverage, Transit was initially chosen to provide these position fixes. Transit is a satellite based navigation system, which was originally developed for the U.S. Navy Polaris submarine fleet by the Applied Physics Lab of Johns Hopkins University. The system has been operational since 1964, and was released for public use in 1967. Reference [10] gives a more complete description.

The Transit Satellite positioning system basically consists of 5 or 6 satellites in low circular polar orbits, transmitting continuously at two very stable frequencies. An earthbound receiver can obtain a position fix whenever a satellite passes overhead, by measuring the Doppler frequency shifts due to the relative motion. The transmitted signals are modulated with a data message containing timing marks and parameters describing the satellites orbit with enough precision to allow the receiver to accurately calculate the absolute position and velocity of the satellite. From this known satellite position and velocity profile and the Doppler derived relative velocity, the receiver can calculate its own position. However the receiver must either be stationary during the satellite pass, as is the case with survey instruments, or the receiver motion (velocity) relative to the earth during the pass must be known (to remove the effect of this velocity on the Doppler measurement). The receiver therefore must be continuously given its velocity. Any error in this velocity input will lead to an error in the position fix that the receiver produces.

Besides requiring velocity inputs, another inconvenient aspect of the Transit system is the waiting time between position fixes. This waiting time can be as short as 15 minutes or as long as 7 hours. Since the Transit satellites are in polar orbits, the mean time between fixes is shorter near the poles than at lower latitudes. In the Canadian arctic (latitude greater than 60°), Transit fixes should occur on average about once every 30 to 50 minutes. This is not exact because the Transit satellite constellation geometry (orbital spacing etc.) is not kept strictly constant.

The accuracy of a Transit position fix is highly variable, depending on many factors, as is briefly explained here. There are two basic types of Transit position errors: Static and Dynamic. The dynamic errors are caused by errors in the velocity and height information that is fed into the Transit receiver by the user, and as such can be considered to be system errors rather than just Transit errors. These dynamic errors are very important for system performance since there is no limit to their size, unless a limit can be placed on the size of the velocity and height errors. Fortunately it was possible to determine the exact relationship between velocity/height input errors and the resulting latitude/longitude output errors, the details and usefulness of which shall be described below.

The static errors are errors which will occur even if the receiver is stationary. These are fairly small but practically unavoidable, and are also described below.

STATIC ERRORS:

There are various sources of static error that are not sufficiently deterministic to completely predict and compensate for [10]:

- Ionospheric Refraction
- Tropospheric Refraction
- Gravitational Field Irregularities

- Drag and Radiation Pressure
- Clock Error
- Oscillator Phase Jitter
- Ephemeris Rounding Error
- Irregularities in the Earth's Motion

Ionospheric refraction introduces an unwanted increase in phase velocity. This results in a position error of about 90 metres in single channel receivers. Fortunately this can be largely compensated for in dual channel receivers by using the two broadcast frequencies (150 MHz and 400 MHz). Since the ionospheric wavelength stretch varies roughly quadratically with the broadcast wavelength, whereas the Doppler shift is linear with frequency, these effects can be separated. The remaining refraction induced error, after compensation, is typically 1 to 5 metres.

Tropospheric refraction also introduces errors, but these are directly proportional to the frequency and thus cannot be eliminated in this way. As with many other Transit errors, the expected size of this error depends strongly on the maximum elevation angle of the satellite, (the angle from the horizon to the satellite, as seen at the receiver position) during the satellite pass. This expected tropospheric refraction error is a function of maximum elevation angle, and is only significant (greater than 15 m) when this angle is less than 10°.

Further Transit position errors result from errors in the geopotential (gravity) model and the surface force model (drag, radiation pressure) used to generate the satellite orbit. These position errors are each on the order of 10 to 30 metres.

There are other less significant but nevertheless identifiable static Transit errors, such as satellite clock error, oscillator jitter, ephemeris rounding error and unmodelled polar motion. These are all in the 1 to 5 metre range, hence not significant in our application.

In summary, the total static error of the Transit position fixes will be uncorrelated in time, and therefore modelled as white measurement noise. The expected magnitude will depend upon the satellite elevation angle and whether the receiver is single or dual channel. For moderate elevation angles (10° to 70°) the dual channel fixes will have rms position errors of about 50 metres, and single channel fixes will have errors of about 100 metres.

DYNAMIC ERRORS:

Reference [11] describes the exact relationship between the velocity error input and the position error output. This relationship is expressed in the form of a sensitivity matrix h , where:

$$\begin{bmatrix} \text{north position error} \\ \text{east position error} \end{bmatrix} = \begin{bmatrix} h_{11} & h_{12} \\ h_{21} & h_{22} \end{bmatrix} * \begin{bmatrix} \text{north velocity error} \\ \text{east velocity error} \end{bmatrix} \quad (21)$$

As is shown in [11], the components of this sensitivity matrix h are complicated functions of the satellite maximum elevation angle, the satellite direction of travel (north to south or vice versa), the receiver latitude, and whether the satellite subpoint is east or west of the receiver at maximum elevation. These are not easily expressed in closed form, but an algorithm to evaluate them has been implemented in the PLANS processor. Figure 5 illustrates the magnitude of the north and east components of the position error (DN, DE), induced by a 1 metre per second velocity error in the north and east directions. These position errors are shown as functions of satellite elevation angle, for a particular receiver latitude and satellite direction of travel. From this we can see that the commonly published relationships (for example [10]) are not valid at high latitudes.

The important fact to be drawn from equation (21) is that it is deterministic and linear in the velocity error. Therefore if there is more uncertainty in the velocity than in the position, equation (21) can be inverted to solve for the velocity error. The Kalman filter effectively does this automatically in PLANS, thereby limiting the velocity uncertainty which otherwise would be unlimited because of the heading error in directional gyro mode, especially near the magnetic pole.

An error in the height supplied to the receiver also produces a Transit position error that can be defined as a function of the same parameters used in the h matrix above. This is also derived in reference [11], but it can be more easily expressed in closed form. The north and east position error dN and dE (true position - Transit position), due to a height error dH (true height - height estimate), can be expressed as

$$\begin{bmatrix} dN \\ dE \end{bmatrix} = \begin{bmatrix} \cos(Y)\tan(\sigma) \\ \sin(Y)\tan(\sigma) \end{bmatrix} dH \quad (22)$$

where σ is the satellite maximum elevation angle, which the receiver supplies, and Y is the bearing from the receiver to the satellite subpoint (at closest approach) which can be computed from σ , the direction of travel, and the receiver position. Equation (22) is derived in reference [11]. The magnitude of this error is illustrated in Figure 6, which shows the position error induced by a 100 metre height error, for a particular receiver latitude and satellite direction of travel and for all satellite elevation angles.

Fig. 5. Velocity Induced Transit Errors

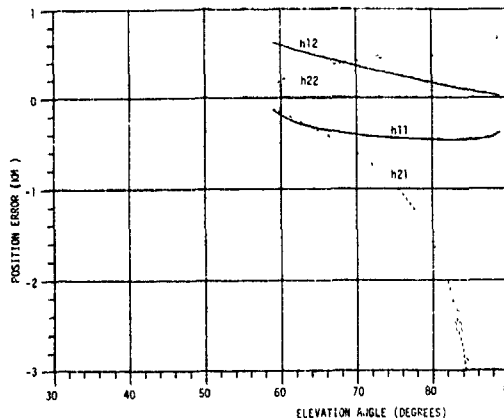
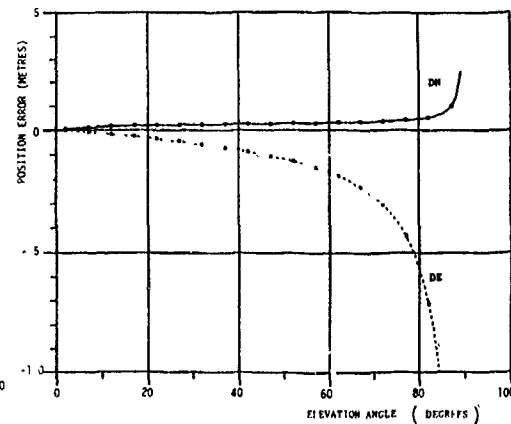


Fig. 6: Height Induced Transit Errors



It is important to notice that, like equation (21), equation (22) is also deterministic and is linear in the height error. The errors described by equations (21) and (22) are in fact independent of each other and of the static errors. These errors are all additive, so that the total Transit position fix error can be expressed as:

$$\begin{bmatrix} dN \\ dE \end{bmatrix} = \begin{bmatrix} \cos(\Psi)\tan(\sigma) & h11 & h12 \\ \sin(\Psi)\tan(\sigma) & h21 & h22 \end{bmatrix} \begin{bmatrix} dH \\ dVn \\ dVe \end{bmatrix} + \begin{bmatrix} s1 \\ s2 \end{bmatrix} \quad (23)$$

where dVn and dVe are the velocity error components and s1 and s2 are the static position error components.

4.5 GPS ERRORS

The Global Positioning System (GPS) is a satellite based navigation system which at time of writing has not been fully implemented. It should be fully operational by about 1990. It is presently scheduled to consist of 18 satellites with 3 hot spares, in 6 orbits at a height of 20,200 Km., with orbital inclination of 55° and period of 12 hours. There will also be a ground control segment and two basic types of receiver equipment, precise or "P-code" receivers for military use, and coarse acquisition or "C/A-code" receivers for civilian or low cost applications. Each satellite will continually broadcast coded messages at two frequencies (1,227.6 MHz and 1,575.4 MHz). The P-code receivers will be able to decode both messages, whereas the C/A-code receivers will only be able to decode one. This system is designed to provide highly accurate 3 dimensional position, 3 dimensional velocity, and time measurements, continuously, on a worldwide basis, with high dynamic capability. The P-code accuracy will be about 15 metres in position (SEP), .1 metre/second in velocity and .1 microsecond in time.

The existing satellite constellation presently provides about 5 to 6 hours of coverage per day in the Ottawa area, in two segments 12 hours apart. This varies with receiver location, and will be altered (expanded) as more satellites are launched and their orbits adjusted. Even these two short periods provide the PLANS system with a significant improvement by bounding the accumulated position and velocity error and by calibrating the drift rates.

The GPS errors are quite small compared to the other PLANS sensor errors, and are largely uncorrelated. Therefore it is not necessary to provide a detailed stochastic model for these errors. They can be simply treated as additive white noise.

5. KALMAN FILTER

It is primarily the errors of the dead reckoning system that the Kalman filter estimates. That is, the north and east position error (in metres), the VRU heading error (in radians), the vehicle odometer scale factor error (dimensionless) and the baroaltimeter height error (metres). In directional gyro mode, the gyro drift rate is also modelled. The Kalman filter also estimates the correlated errors of the aiding sensors, such as the magnetic flux valve heading error (radians) and the error of the digital map height (metres).

The filter estimates these errors by processing all of the available measurements using suitable stochastic and deterministic error models. Reference [12] describes the standard stochastic model types and provides an elementary introduction to Kalman filtering techniques.

5.1 STATE VECTOR

The PLANS filter employs an eight dimensional state space. The eight states are defined as follows:

$$\begin{array}{l|l}
 \begin{array}{l}
 X1 \\
 X2 \\
 X3 \\
 X4 \\
 X5 \\
 X6 \\
 X7 \\
 X8
 \end{array} &
 \begin{array}{l}
 \text{true height above geoid - map height} \\
 \text{true height above geoid - baroaltimeter height} \\
 \text{true heading - magnetic heading} \\
 \text{gyro drift rate in directional gyro mode} \\
 \text{true heading - gyro heading} \\
 (\text{true speed - odometer speed})/\text{odometer speed} \\
 (\text{true latitude - DR latitude}) \text{ in metres} \\
 (\text{true longitude - DR longitude}) \text{ in metres}
 \end{array}
 \end{array} \quad (24)$$

It should be clarified that these states represent the Markov, or time correlated portions of the errors. The uncorrelated component of the sensor errors are treated as measurement noise. Also there are actually two state vectors involved, because state X4 is deleted when in gyrocompass mode.

5.2 TRANSITION MATRIX

DR position error is simply the integral of the DR velocity error. This fact, together with the assumed stochastic error models for the DR velocity errors and the remaining state vector components, is used to propagate the state vector in time, according to the vector differential equation:

$$\dot{X}(t) = F(t)X(t) + V(t) \quad (25)$$

This is used to propagate X between measurements. To implement this on a digital computer, equation (25) was converted to a discrete difference equation and linearized, to obtain the STATE TRANSITION matrix Φ , where:

$$X(t+\Delta t) = \Phi X(t) + V(\Delta t) \quad (26)$$

The result is

$$\begin{array}{c|c|c|c|c|c}
 \begin{array}{l}
 X1 \\
 X2 \\
 X3 \\
 X4 \\
 X5 \\
 X6 \\
 X7 \\
 X8
 \end{array} &
 \begin{array}{c}
 -\Delta t/T1 \\
 e \\
 -\Delta t/T2 \\
 e \\
 -\Delta t/T3 \\
 e \\
 \boxed{} \\
 -\Delta t/T6 \\
 e \\
 \Phi_{7,5} \quad \Phi_{7,6} \quad 1 \quad 0 \\
 \Phi_{8,5} \quad \Phi_{8,6} \quad 0 \quad 1
 \end{array} &
 \begin{array}{c}
 X1 \\
 X2 \\
 X3 \\
 X4 \\
 X5 \\
 X6 \\
 X7 \\
 X8
 \end{array} &
 \begin{array}{c}
 V1 \\
 V2 \\
 V3 \\
 V4 \\
 V5 \\
 V6 \\
 0 \\
 0
 \end{array}
 \end{array} \quad (27)$$

where:

$$\begin{aligned}
 \Phi_{7,5} &= -\Delta t S \sin(\theta + X5/2) \\
 \Phi_{7,6} &= \Delta t S \cos(\theta + X5) \\
 \Phi_{8,5} &= \Delta t S \cos(\theta + X5/2) \\
 \Phi_{8,6} &= \Delta t S \sin(\theta + X5)
 \end{aligned} \quad (28)$$

where Δt is the propagation time interval, S is the deadreckoning speed, θ is the deadreckoning heading, and the submatrix

$$\begin{array}{c|c|c}
 \boxed{} &
 \begin{array}{c}
 -\Delta t/T4 \\
 e \\
 -\Delta t/T4 \\
 T4(1-e) \quad 1
 \end{array} &
 \begin{array}{l}
 \text{in directional gyro mode} \\
 \\
 \text{in gyrocompass mode}
 \end{array}
 \end{array} \quad (29)$$

where:

$$\begin{aligned}
 T1 &= \text{map height error correlation time} \\
 T2 &= \text{baroaltimeter height error correlation time} \\
 T3 &= \text{magnetic declination error correlation time} \\
 T4 &= \text{gyro drift rate correlation time} \\
 T5 &= \text{gyro Markov error correlation time} \\
 T6 &= \text{odometer speed factor error correlation time}
 \end{aligned} \tag{30}$$

and

$$W_i = \text{driving white noise with covariance } Q_i \text{ (for } i = 1 \text{ to } 6) \tag{31}$$

where (see [12]):

$$Q_i = M_i (1 - e^{-2\Delta t/T_i}) \tag{32}$$

where the M_i are the Markov process steady state root mean squared values:

$$\begin{aligned}
 M1 &= \text{map height error covariance} \\
 M2 &= \text{baroaltimeter height error covariance} \\
 M3 &= \text{magnetic declination error covariance} \\
 M4 &= \text{gyro drift rate covariance} \\
 M5 &= \text{gyro heading error covariance} \\
 M6 &= \text{speed factor error covariance}
 \end{aligned} \tag{33}$$

Notice from equations (27) and (28) that the state transition equation has not been completely linearized in X , and consequently the transition matrix Φ still exhibits a weak dependence on the state vector X . Therefore this PLANS design uses an extended Kalman filter rather than a simple linear one. This is necessary because of the potentially large deadreckoning heading error, $X5$, that could occur in directional gyro mode, and the fact that the DR position error, $(X7, X8)$, varies nonlinearly with this heading error. In fact the linearization is only necessary to propagate the covariance matrix. The state vector components, and in particular $X7$ and $X8$, are propagated in a nonlinear manner, and at a higher rate than the covariance.

This nonlinearity in heading and potentially large heading error also led to a closed loop filter design whereby the filtered estimate of velocity and height are fed into the Transit receiver. Periodic resetting of the VRU heading with the filter heading keeps the VRU error $X5$ bounded. This is important primarily for the purpose of using DR as a backup under certain failure conditions.

5.3 MEASUREMENT VECTOR

The odometer "speed" and gyro heading are measured at a 1 Hz rate for dead reckoning, and the digital map is read once a minute to obtain the height. The filter measures the baroaltimeter and the magnetic flux valve once a minute to update its estimate of the state vector X (especially $X1$, $X2$ and $X3$). Transit position fixes are processed whenever they occur to update the filter estimate of X (especially $X7$ and $X8$ but also $X5$ and $X6$). Whenever the vehicle is in directional gyro mode and is not moving, the gyro drift rate is measured from successive heading measurements, to allow the filter to update the estimate of $X4$. The GPS position, velocity and height are measured at a 30 second rate, when they are available. The Kalman filter's measurement vector is therefore defined as:

Z1	map height	-	barometric height	(34)
Z2	map height	-	GPS height	
Z3	magnetic heading	-	gyro heading	
Z4	(gyro heading(t+Δt) - gyro heading(t))/Δt	-		
Z5	Transit latitude	-	DR latitude (in metres)	
Z6	Transit longitude	-	DR longitude (in metres)	
Z7	GPS latitude	-	DR latitude (in metres)	
Z8	GPS longitude	-	DR longitude (in metres)	
Z9	GPS north velocity	-	DR north velocity (m/sec)	
Z10	GPS east velocity	-	DR east velocity (m/sec)	

The Transit position fixes occur at irregular intervals, averaging about once every 90 minutes at low latitudes and more frequently at higher latitudes.

As was mentioned, in order to be at all useful at high latitudes, the raw magnetic flux valve measurement must first be adjusted using a magnetic declination value. The PLANS software generates this magnetic declination using the IGRF85 geomagnetic field model, described in reference [7]. This field model predicts all three components of the earth's magnetic field, which allows not only the magnetic declination to be calculated, but also the horizontal field strength and the dip angle, which can be used to give a reasonable indication of the accuracy that can be expected from the flux valve. The magnetic heading measurements are also adjusted using the calibration function described in reference [1] which corrects for the permanent and induced fields of the vehicle, and the sensor misalignment.

5.4 MEASUREMENT MATRIX

The relationship between the state vector X and the measurement vector Z must be described in detail for the Kalman filter to properly process the measurements. This description is expressed as a measurement equation:

$$Z = H \cdot X + V \quad (35)$$

where H is the "measurement matrix" and V is the measurement noise vector. For PLANS the measurement equation is:

$$Z = \begin{bmatrix} -1 & 1 & & & & & & & & \\ -1 & 0 & & & & & & & & \\ & & -1 & 0 & 1 & & & & & \\ & & 0 & -1 & 0 & & & & & \\ H51 & 0 & 0 & 0 & H55 & H56 & 1 & 0 & & \\ H61 & 0 & 0 & 0 & H65 & H66 & 0 & 1 & & \\ & & & & & & 1 & 0 & & \\ & & & & & & 0 & 1 & & \\ & & & & & & & & -S \sin \theta & S \cos \theta \\ & & & & & & & & S \cos \theta & S \sin \theta \end{bmatrix} \begin{bmatrix} X1 \\ X2 \\ X3 \\ X4 \\ X5 \\ X6 \\ X7 \\ X8 \\ V1 \\ V2 \\ V3 \\ V4 \\ V5 \\ V6 \\ V7 \\ V8 \\ V9 \\ V10 \end{bmatrix} + \begin{bmatrix} V1 \\ V2 \\ V3 \\ V4 \\ V5 \\ V6 \\ V7 \\ V8 \\ V9 \\ V10 \end{bmatrix} \quad (36)$$

where:

$$\begin{aligned} S &= \text{vehicle speed} \\ \theta &= \text{vehicle heading} \\ H51 &= \cos(Y) \tan(\sigma) \\ H61 &= \sin(Y) \tan(\sigma) \\ Y &= \text{bearing to subpoint} \\ \sigma &= \text{maximum elevation angle} \end{aligned} \quad (37)$$

The remaining four H components ($H55$, $H56$, $H65$ and $H66$), which define the sensitivity of the Transit fix to errors in velocity, are described briefly in reference [1] and are related to the h matrix in equation 21 by a simple transformation:

$$\begin{bmatrix} H55 & H56 \\ H65 & H66 \end{bmatrix} = \begin{bmatrix} h11 & h12 \\ h21 & h22 \end{bmatrix} * \begin{bmatrix} -S \sin \theta & S \cos \theta \\ S \cos \theta & S \sin \theta \end{bmatrix} \quad (38)$$

The measurement noise vector V is defined as the non-Markov component of the measurement vector Z , given by equation (34). These are assumed to be uncorrelated white noise processes. The covariance matrix R of this noise vector is therefore diagonal of rank 10.

$$R = E\{VV^T\} = \begin{bmatrix} R1 & & & & & & & & & \\ & R2 & & & & & & & & \\ & & R3 & & & & & & & \\ & & & R4 & & & & & & \\ & & & & R5 & & & & & \\ & & & & & R6 & & & & \\ & & & & & & R7 & & & \\ & & & & & & & R8 & & \\ & & & & & & & & R9 & \\ & & & & & & & & & R10 \end{bmatrix} \quad (39)$$

where the numerical values of $R1$ $i=1,8$ are all constants except for $R2$, $R3$, $R7$ and $R8$.

The GPS vertical, north and east noise $R7$, $R7$ and $R8$ are modelled simply as constants times the appropriate geometric dilution of precision (GDOP). These GDOPs are supplied by the receiver.

To estimate the magnetic heading noise covariance $R3$ we consider separately the "static" errors R_s caused by the random field fluctuations and the dynamically induced errors R_d , described by equations (11) and (12) respectively. Thus we let

$$R3 = R_s^2 + R_d^2 \quad (40)$$

where equation (11) implies that the level of the static noise R_s (since we don't know what the field errors dX and dY are) is inversely proportional to the horizontal magnetic field strength. Of course a

heading noise level of greater than π radians (180°) is physically meaningless, so it is logical to express the static heading noise level (square root covariance in radians) in the form:

$$R_s = \frac{FL}{(H + FL/\pi)\sqrt{N}} \quad \text{radians} \quad (41)$$

where H is the local horizontal field strength in gammas, F is the low latitude field strength in gammas, (about 60,000 γ) L is the low latitude magnetic heading noise level in radians and N is the number of samples averaged in the prefilter. The $1/\sqrt{N}$ reduction in noise is due to the prefilter averaging, assuming that the measurement noise is independent and less than π . Equation (41) will have the correct asymptotic values and approximately the correct H dependence. At low latitudes the magnetic noise FL is about 100 γ , resulting in a low latitude magnetic heading noise R_s of about .0018 radians (.1°). Of course at higher latitudes this increases considerably, for example at 65° latitude -100° longitude H is about 4,600 γ , resulting in a magnetic heading noise of about .02 radians (1.2°).

Equation (12) shows that the dynamically induced gimbal error produces a heading error that is also proportional to $1/H$. Although the gimbal attitude and accelerations are unknown, since they are in response to the vehicle moving over rough terrain it is reasonable to approximate their effect as being proportional to the velocity:

$$R_d = \frac{CS}{H\sqrt{N}} \quad (42)$$

where S is the speed and C is an experimentally determined constant.

5.5 FILTER IMPLEMENTATION

The filter mechanism used is Biermans UDU [13], which offers both excellent numerical stability (preserving a positive definite symmetric covariance) and reasonably good computational efficiency. Since the state vector is quite small no special measures are required to reduce the processing burden.

Special measures did have to be taken to handle certain practical problems that are expected to arise. One such difficulty is that it may be necessary to start the system in directional gyro mode without an adequate estimate of the initial heading. Because of this possibility it was necessary to implement a "heading adjustment" parameter, to be added to the directional gyro measurements before they are passed to the navigation filter. The value of this parameter will be initially determined by using the magnetic flux measurement. This value will then be refined at an hourly rate by the Kalman filter itself. This is necessary whenever directional gyro mode is used for any substantial length of time (10 hours or so) because the growing directional gyro heading error will introduce serious nonlinearities into the filter equations. For example the linearized state transition equation (27) carries the implicit assumption that X_5 (the gyro heading error) is small.

5.6 PREFILTER

To be as accurate and reliable as possible, any Kalman filter based integrated system should have a good prefilter to apply any known corrections to the measurements, to remove any spurious data and to detect sensor or subsystem failures.

To maximise the DR accuracy the VRU must be carefully aligned to the vehicles direction of motion, and the odometer scale factor must be carefully calibrated. The results of this alignment and calibration are used by the prefilter to offset the VRU measurement and scale the odometer measurements appropriately.

The magnetometer measurement would be useless without the geomagnetic field model, which allows the prefilter to convert from magnetic to geographic or true heading. The prefilter must also apply the magnetic calibration function for the vehicles permanent and induced fields, which is a function of the true and magnetic headings. Even this corrected and calibrated magnetic heading is very noisy. By averaging N magnetic measurements in the prefilter, the uncorrelated portion of this noise can be reduced by $1/\sqrt{N}$. Of course this can only be done over a short time interval over which the correlated error will not change appreciably, and it is the (gyro - magnetic) heading that is averaged, to remove the effect of heading change.

Of course the prefilter must also convert the pressure measurement to height, using the standard atmosphere model, and the prefilter must form the map height measurement by reading grid heights from the database and interpolating.

The prefilter also uses various methods to detect spurious measurements and hard sensor failures, using known physical constraints on the vehicles speed, acceleration, turn rate etc. All measurements are first tested in this simple way, and if they pass this, they then undergo a residual test [14]. The residual test uses the Kalman filters estimate of what the measurement error should be, based on all previous measurements, along with the filters covariance (degree of uncertainty in the estimate) based on the stochastic error models. This residual test is the most sensitive test, and the last line of defence before the measurement is processed by the filter. These two levels of failure detection significantly improve the accuracy and reliability of the system.

6 SIMULATIONS

Throughout the PLANS development activity at DREO, extensive simulations have been performed for three important reasons:

- 1/ system design testing and evaluation
- 2/ algorithm and software testing and evaluation
- 3/ system performance prediction

Extensive Monte-Carlo simulations were performed. These simulations predict the PLANS performance under conditions that could not be obtained outside the arctic. Simulating high latitude conditions and very long excursions is fairly easy, fast and cost effective, whereas an arctic trial is very time consuming and expensive. The simulation runs are also very useful for comparing the performance of many different algorithm designs under identical conditions.

One primary aim of the system design testing was to determine whether or not the Kalman filter could correct the velocity errors (especially the heading error) in the region where both magnetic and gyroscopic heading is poor, by using the dynamic Transit position fixes.

Simulations were performed using the full PLANS software package, along with specially created measurement generation routines. Monte-Carlo and Covariance Analysis techniques were employed to predict the PLANS system performance. A very brief summary of some of the basic simulation results is presented here in the form of two Monte-Carlo sets. These were run without GPS since the system will be essentially slaved to GPS when it is available, and before the VRU resets had been implemented. For the first set, the simulated starting position was 65° N latitude and 100° W longitude, and for the second it was 75° N 100° W, which is very close to the geomagnetic pole. The first set was run with a constant simulated speed of 10 metres per second, and the second with a periodically varying speed profile that changed smoothly between 0 and 10 metres per second, dwelling several hours at each extreme. This was to show the effect of the gyro drift rate measurements, and stationary Transit fixes. In both runs the simulated heading was 45° (northeast), and the duration was 24 hours. The second simulation therefore traverses the area roughly between the magnetic pole and the northern limit of land, providing a worst case scenario.

Each Monte-Carlo set consisted of 20 individual simulated runs of 24 hours duration each. In each of the 20 runs of a given set the deterministic errors were exactly the same and the random errors had the same statistical properties (root mean squared values and correlation times) but different actual values. The results of each set were combined to produce one representational run. Figures 7 and 8 illustrate the results of the first set, and the numerical results below summarize the basic accuracy results.

PLANS FMS POSITION ERROR	282 metres
TRANSIT RMS POSITION ERROR	303 metres
DR RMS POSITION ERROR	21,426 metres
PLANS RMS HEADING ERROR	.67 degrees
DIRECTIONAL GYRO RMS HEADING ERROR	4.5 degrees
FLUX VALVE RMS HEADING ERROR	2.1 degrees

The solid curve in Figure 7 shows the radial position error of the PLANS system over the 24 hour simulation, starting at 65° latitude, -100° longitude, and heading north east with a constant speed of 10 m/s. As in each of the simulation plots, each point is actually the root-mean-square of 20 points; one from each simulation run. (This is the basis of Monte-Carlo simulation, and produces a smoother, more statistically significant plot.) The dashed curve in Figure 7 is the Kalman filter's estimate of its radial position accuracy from the covariance matrix, representing the 95% probability level (about 2σ). Figure 8 shows the PLANS heading error (solid curve) and the filter's 68% (1σ) error estimate (dashed). The overall accuracy results tabulated above are found by taking the rms of each of these curves over the 24 hours.

Fig. 7.
Simulated PLANS Position Error and Covariance

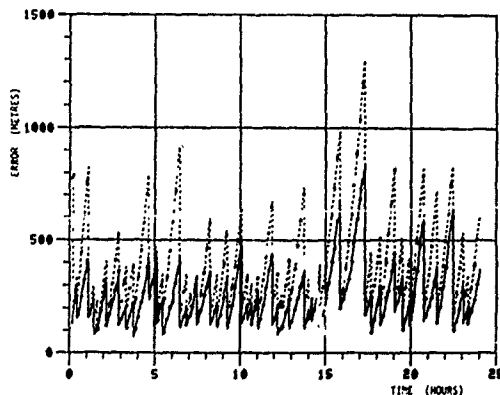
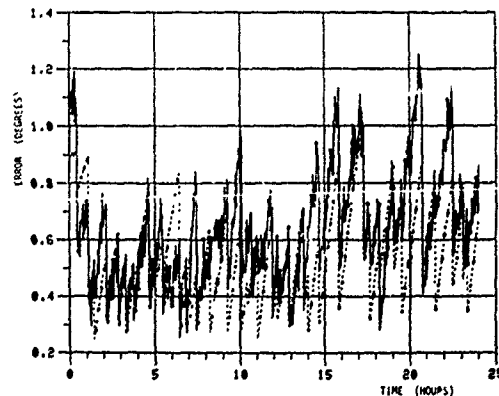


Fig. 8.
Simulated PLANS Heading Error and Covariance



In directional gyro mode, the heading error is expected to increase without bound resulting in an increasingly rapid growth in the DR position error and, more importantly, a potential for unbounded growth in the Transit position error (if the Kalman filter can not adequately estimate and remove the gyro error). Here it can be seen that the filter performed well, keeping the PLANS heading error, and hence the position error, well bounded.

The basic results of the second Monte-Carlo set, with the periodic velocity profile at the higher latitude, northeast from the geomagnetic pole, are:

PLANS RMS POSITION ERROR	350 metres
TRANSIT RMS POSITION ERROR	326 metres
DR RMS POSITION ERROR	10,640 metres
PLANS RMS HEADING ERROR	.96 degrees
DIRECTIONAL GYRO RMS HEADING ERROR	6.4 degrees
FLUX VALVE RMS HEADING ERROR	10.7 degrees

This, along with other simulations, verifies the potential effectiveness of the Transit position fixes as a source of heading information, and demonstrates proper blending of the VRU and magnetic heading.

Sensitivity analysis simulations were also performed, by running various Monte-Carlo sets, each with a significantly increased value being assigned for a major error source. As would be expected the position error was found to be most sensitive to the VRU error.

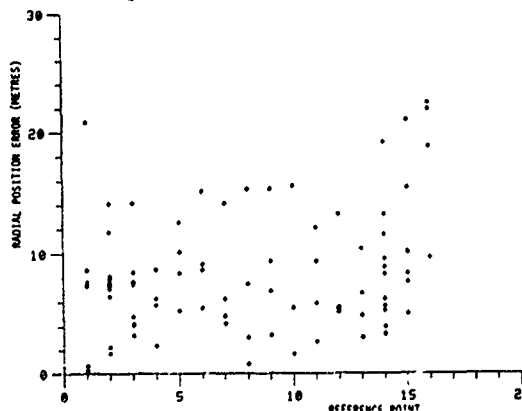
7. FIELD TRIALS

Three types of PLANS trials have been conducted to date: local trials conducted at DREO on smooth level roads at moderate speeds, and formal field trials conducted both on a military base, over a variety of terrain, at maximum speed and in the arctic under winter conditions. Naturally the results are very different. The platform for all of these trials was an M577 armoured tracked vehicle.

7.1 LOCAL TRIALS

The local trials were conducted periodically from early 1985 to late 1986 on a course of about 8 kilometres. The primary purpose of the local road trials was to determine the error model parameters for the various sensors and subsystems, to verify the structure of these error models, and to evaluate different brands of each type of sensor for final sensor selection. Of course debugging of software and interface problems, and the calibration and alignment of the heading sensors was also done at this time. Another goal of these trials was to verify the simulation results, and gain confidence in the error models. By the fall of 1986 these objectives had all been met. The position error never exceeded 100 metres, and typical RMS position error over the course was about 15 metres. Figure 9 shows the radial errors from one week of local trials, as a function of reference points, which were on average about 500 metres apart. The high accuracy was due to the shortness of the course, and the hard smooth road.

Fig. 9. PLANS Road Trial Results, 1985.



7.2 SOUTHERN FIELD TRIAL

The first formal field trial was conducted on a course about 55 kilometres in length, over a one week period in September 1986. The purpose of this trial was to evaluate the system performance, and also to provide a significant database of raw measurements to continue to enhance and refine the system software. Recording all of the raw measurements as well as the reference data allows the trial to be "rerun" back at the lab, providing a powerful tool for further system development and detailed sensor error analysis.

This field trial was designed to generate the largest position error that could be expected for the location (about 46° N, 77° W). This was accomplished by driving at full speed for over half an hour (the expected time between Transit fixes) in a direction generally away from the starting point, and returning

along a different path, again at full speed. The test plan was to conduct a series of these "runs", each of which consists of one complete circuit of the 55 km. course. Each run took about 100 minutes. This was repeated as often as time would allow, over four days. The PLANS position, velocity, heading etc. as well as all of the raw sensor data and the reference data, was recorded on magnetic tape for later analysis.

The reference position information, for comparison to the PLANS data, was obtained by placing 27 simple retro-reflectors at surveyed positions at approximately 2000 metre intervals along the course. An infrared emitter/detector was mounted on the vehicle, so that as the vehicle drove past each surveyed point a reflection would be detected and recorded with timing information, in such a way that the surveyed position could be later compared to the PLANS position, and the PLANS position accuracy could therefore be determined.

The real time PLANS position and heading accuracy results are summarized in Table I. Although there were 27 reflectors on the course, seldom was a good reflection received from every one, so Table I gives the number of data points (good reflections) obtained for each run.

The behaviour of PLANS position error around the course (ie. in time) is illustrated in Figure 10 where the RMS radial position error is plotted as a function of the reference points. From this it can be seen that the position error was just under 2% of distance travelled, to a maximum of about 600 metres. The position and heading errors shown in Table I and Figure 10 are larger than was expected. The fact that the heading error averaged over all runs was not small compared to the RMS heading error indicates clearly that there was a significant bias in the heading error, due to an inaccurate installation alignment. This misalignment was determined to be about 1.7° , and was a result of having to change vehicles a few days before the trial.

The effect of correct installation can be simulated by running the data through the PLANS software with a constant 1.7° added to all raw VRU heading data to compensate for the estimated installation misalignment. This is equivalent to having rotated the VRU by 1.7° on installation, as should have been done. With this constant added to the VRU measurements, and all other measurements exactly as originally collected, the PLANS software was rerun, and the results are shown in Table II.

TABLE I: FIELD TRIAL RESULTS

RUN #	NUMBER OF		POSITION ERROR RMS (metres)	HEADING ERRORS	
	DATA POINTS	TRANSIT FIXES		AVERAGE (degrees)	RMS (degrees)
1	27	0	335	1.1	1.4
2	17	0	143	-1.1	1.5
3	25	1	608	3.1	3.3
4	27	1	404	1.9	2.2
5	23	0	411	1.6	2.1
6	16	1	270	0.6	1.2
7	19	0	421	1.4	1.6
8	25	1	327	0.7	2.2
9	13	0	141	0.4	1.0
10	17	1	379	-0.5	2.7
11	15	0	494	0.9	2.1
Total:	224	5	392	1.2	2.1

TABLE II: RESULTS WITH REALIGNED VRU

RUN #	NUMBER OF		POSITION ERROR RMS (metres)	HEADING	
	DATA POINTS	TRANSIT FIXES		ERROR RMS (degrees)	RMS (degrees)
1	27	0	101	.8	
2	17	0	339	2.2	
3	25	1	135	.7	
4	27	1	43	.4	
5	23	0	186	1.2	
6	16	1	107	.5	
7	19	0	66	.5	
8	25	1	251	2.4	
9	13	0	324	1.9	
10	17	1	315	2.1	
11	15	0	146	1.3	
Total:	224	5	198	1.4	

As can be seen from Table II, this improvement in installation alignment accuracy leads to a very significant improvement in PLANS performance. Figure 11 shows the RMS radial position error as function of reference point, which can be compared to the original (badly aligned) results shown in Figure 10. The position error here is now less than 1% of distance travelled, to a maximum of about 300 metres.

Fig. 10. Field Trial Results (Not Realigned)

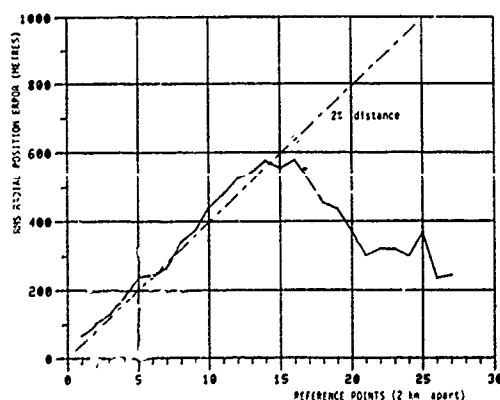


Fig. 11. Field Trial Results (Realigned)

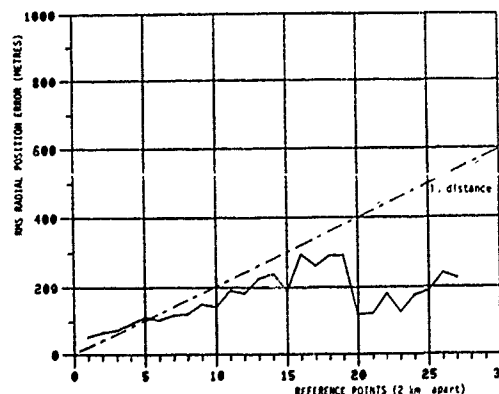
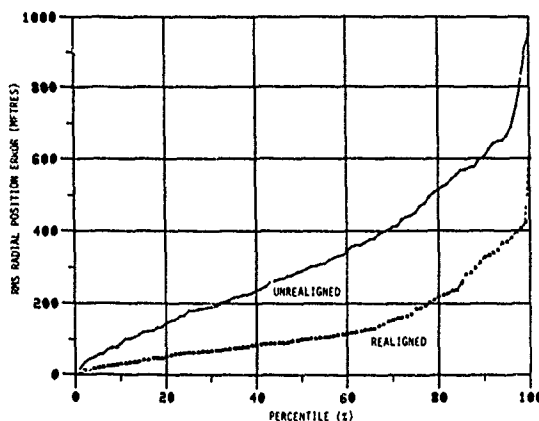


Figure 12 shows the statistical distribution of these radial position errors in the form of percentiles. For example the 50 percentile (the CEP) is seen to be about 290 metres originally, and 100 metres with the realignment. The 68 percentile (1 σ) is about 400 metres originally, and 150 metres realigned. The smoothness of these percentile plots, and the agreement between the 68 percentiles and the RMSs, is evidence that the data set is large enough to yield statistically significant results.

Fig. 12. Field Trial Percentile Results, November 1986.



One obvious but valuable lesson from this trial is the importance of the initial gyrocompassing accuracy of the VRU. This effects the installation alignment and the runtime initialisation. One way to reduce the initial gyrocompassing error is by cycling through the initial gyrocompass phase several times. If the settling errors are uncorrelated and the alignment time is not excessive then N alignments can be performed consecutively (without moving of course) and averaged to reduce the error by a factor of $1/\sqrt{N}$. Four cycles would then cut this error in half.

It should be mentioned that the effect of multiple gyrocompassing at initialization will be offset at higher latitudes by the expected increase in this error. For example at 70° latitude this error will be greater than it is at 45° latitude by a factor of:

$$\frac{\cos(45^\circ)}{\cos(70^\circ)} = 2.07$$

Therefore the RMS heading accuracy from Table II should also be representative of what can be achieved with PLANS at high latitude, if the multiple gyrocompassing is effective in reducing the initial heading error.

7.3 ARCTIC TRIAL

An arctic trial was conducted in February 1987, to test the PLANS equipment under severe weather conditions, at high latitude, over icy terrain. The location selected was Iqaluit, at a latitude of about 64° N. Weather conditions at the time were somewhat worse than normal, with 60 km/hr winds, an ambient temperature of about -40°C and occasional whiteouts. Air transport scheduling for the host vehicle limited the test to three days, one of which was required for preparation. This shortness of time, together with the snowcover, limited the reference data to 4 surveyed points within the town of Iqaluit and one near a radio tower about 5. km. away. In fact none of the survey markers were exposed, however the 4 nearby points could be determined to within a few metres by reference to more prominent nearby structures and the distant point could be estimated to within about 100 m.

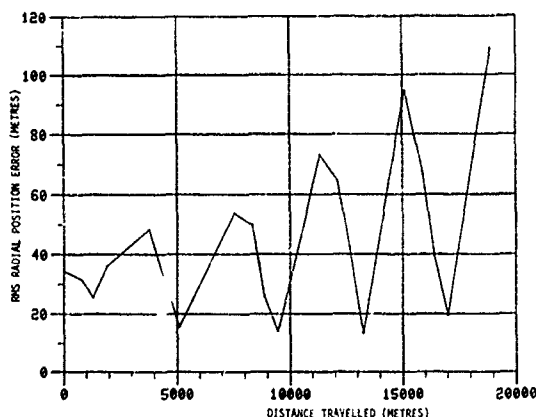
The course for this trial consisted of one closed circuit about 3.7 km. in length around the 4 survey points, and a return trip to the outlying radio tower and back, about 10.3 km. in length. The surface was mostly hard packed snow on gravel or asphalt, with some drifting. The results are summarized in Table III and Figure 13. This data was collected from one long trip (run # 5) and 14 short circuits. Due to an antenna failure, no GPS data was received on this trial. This led to larger than normal vertical errors, which corrupted the Transit position fixes. In fact for these short periods, given an accurate initial position, the system performed better without Transit than with it, as shown in Table III. These results without Transit were obtained by rerunning the data through the filter software without the Transit measurements. The small long term error growth seen in Figure 13 (< .5% distance) is due to the cancellation of heading error effects during the return path. The short term growth is more realistic, but it will be bounded by the satellite fixes.

The subsystems that were expected to have difficulty in arctic conditions, and were therefore of primary interest here, were the gyro, odometer and magnetometer. From these results it is evident that the integration of these sensors is possible under arctic conditions.

TABLE III. ARCTIC TRIAL RESULTS

RUN #	DATA POINTS	TRANSIT FIXES	DURATION OF RUN (min.)	RMS POSITION ERRORS (metres)	
				WITH TRANSIT	WITHOUT TRANSIT
1	21	0	143	45	45
2	13	0	83	37	37
3	5	1	52	62	56
4	17	2	132	82	47
5	3	0	65	85	85
6	3	1	70	177	18
Total	62	4	545	71	47

Fig. 13. Arctic Trial Results, February 1987.



8. CONCLUSIONS

An exploratory development model of an optimally integrated system for arctic land navigation has been designed and implemented and tested at DREO. All necessary sensors, processors and displays have been identified, purchased, tested and interfaced to:

- 1/ Vertical Reference Unit (gyrocompass/directional gyro)
- 2/ Dual Channel Transit Receiver
- 3/ C/A Code GPS Receiver
- 4/ Pickoff for the APC odometer
- 5/ Magnetic Compass Sensor Unit (fluxgate magnetometer)
- 6/ High Output Pressure Transducer
- 7/ Motorola 68000 microprocessor
- 8/ Custom LCD displays

The necessary algorithms and data bases have been acquired or developed to obtain the most accurate possible navigation information from these sensor measurements, including a full geomagnetic field model algorithm and a digital arctic elevation map data base. A detailed functional relationship has been derived which explicitly describes the dynamic Transit position fix errors in such a way as to allow heading information to be extracted from the Transit position fixes.

A two stage prefilter, and an 8 state extended Kalman filter have been designed and implemented to optimally integrate all valid sensor information. Extensive simulations have been conducted to test the PLANS system design with encouraging results. Local road trials have been conducted to test the real-time PLANS software, the sensors and sensor interfaces, and to refine the error models required in the Kalman filter design. A formal field trial was conducted to evaluate the system accuracy performance, with positive results.

In summary an automatic, all weather vehicle navigation system has been developed to the exploratory development model stage, capable of providing the accuracy and reliability necessary for the arctic application.

9. FUTURE DIRECTIONS

An extended high arctic trial is planned, to evaluate the high latitude and cold temperature performance of PLANS. Steps are presently being taken to transfer the PLANS technology to industry in order to have an advanced development model built, with the intention to shortly thereafter have several units produced for immediate use.

As an integrated system PLANS has been designed with sufficient flexibility to allow new sensors to replace old with a minimum of reconfiguration. The PLANS sensor suite has already changed several times since work began, as better sensors have become available. The existing sensor suite is still not entirely satisfactory and some possible replacements are presently being considered. There are also several new sensors, expected to be available in the near future which are good candidates for integration into PLANS. A moderately priced RLG or DTG ABRs would be very desirable. Part of the emerging technology is a GPS receiver that also measures heading and attitude, using several antennae, which could also be very useful.

REFERENCES

- [1] McMillan, J.C., "Design of an Optimally Integrated Primary Land Arctic Navigation System, Vol. I System Design", DREO Report 946, 1986.
- [2] Whitham, K., Loomer, E.I., and Niblett, E.R. "The Latitudinal Distribution of Magnetic Activity in Canada", J. of Geophysical Research, Vol. 65, No. 12, Dec. 1960.
- [3] Gupta, R.R., Donnelly, S.P., Creamer, P.M., and Sayer, S., "Omega Signal Coverage Prediction Diagrams for 10.2 kHz", Document prepared by The Analytic Sciences

Corporation for the U.S. Department of Transportation, U.S. Coast Guard, October 1980.

- [4] Ayers, H., JMR Instruments Canada Ltd., "Azimuth Determination Using Transit Satellite System", July 1983.
- [5] McMillan, J.C., "A Kalman Filter for Marine Navigation", M Phil thesis University of Waterloo, Waterloo, Ontario, Canada, 1980.
- [6] Liang, D.P., McMillan, J.C. and Maskell, C.A., "Design and Test Evaluation of a Marine Integrated Navigation System", Proceedings of the National Technical Meeting, The Institute of Navigation, 1984.
- [7] Malin, S.R.C., and Barraclough, D.R., Computers and Geosciences, Vol. 7 No. 4, 1981, pp. 401-405.
- [8] Kayton, M., and Fried, W.R., editors, "Avionics Navigation Systems", John Wiley & Sons, Inc., Toronto, 1969.
- [9] Anderson, J.D. Jr., "Introduction to Flight", McGraw-Hill, 1978, p. 59.
- [10] Stansell, T.A., "Transit, The Navy Navigation Satellite System", Navigation, Vol. 18, No. 1, Spring 1973.
- [11] McMillan, J.C., DREO report, in progress.
- [12] Gelb, A., Editor, The Analytic Sciences Corporation, "Applied Optimal Estimation", Cambridge, Massachusetts, The M.I.T. Press, 1974.
- [13] Bierman, G.J., "Factorization Methods for Discrete Sequential Estimation", New York, Academic Press, 1977.
- [14] McMillan, J.C., "Optimal Compensation of Marine Navigation Sensor Errors", proceeding of the 7th IFAC/IFORS Symposium on Identification and System Parameter Estimation, York, U.K., July 1985.

REPORT DOCUMENTATION PAGE											
1. Recipient's Reference	2. Originator's Reference AGARD-AG-314	3. Further Reference ISBN 92-835-0566-2	4. Security Classification of Document UNCLASSIFIED								
5. Originator	Advisory Group for Aerospace Research and Development North Atlantic Treaty Organization 7 rue Ancelle, 92200 Neuilly sur Seine, France										
6. Title	ANALYSIS, DESIGN AND SYNTHESIS METHODS FOR GUIDANCE AND CONTROL SYSTEMS										
7. Presented											
8. Author(s)/Editor(s) Various	Edited by Professor C.T.Leondes		9. Date June 1990								
10. Author's/Editor's Address See Flyleaf			11. Pages 514								
12. Distribution Statement	This document is distributed in accordance with AGARD policies and regulations, which are outlined on the Outside Back Covers of all AGARD publications.										
13. Keywords/Descriptors	<table border="0"> <tr> <td>Control equipment</td> <td>Air navigation</td> </tr> <tr> <td>Guidance computers</td> <td>Air traffic control</td> </tr> <tr> <td>Navigational aids</td> <td>Communication equipment</td> </tr> <tr> <td>Flight control</td> <td>Surface navigation</td> </tr> </table>			Control equipment	Air navigation	Guidance computers	Air traffic control	Navigational aids	Communication equipment	Flight control	Surface navigation
Control equipment	Air navigation										
Guidance computers	Air traffic control										
Navigational aids	Communication equipment										
Flight control	Surface navigation										
14. Abstract	<p>The field of modern guidance and control systems has been raised to a very high level of capability because of the powerful high technology advances of the past several decades. This NATO AGARDograph captures the spirit of these powerful capabilities, and it is structured in 8 parts.</p> <p>These 8 parts are</p> <p>Part I — Integrated Guidance and Control Systems ;</p> <p>Part II — NAVSTAR/GPS Systems ;</p> <p>Part III — Optical Gyroscope and Control Systems ;</p> <p>Part IV — Integrated Communication and Navigation Systems ;</p> <p>Part V — Integrated Navigation/Flight Control Systems ;</p> <p>Part VI — Civil Aircraft Navigation and Traffic Control ;</p> <p>Part VII — Special Topics</p> <p>Part VIII — Land Navigation Systems .</p>										

<p>AGARDograph No.314 Advisory Group for Aerospace Research and Development, NATO ANALYSIS, DESIGN AND SYNTHESIS METHODS FOR GUIDANCE AND CONTROL SYSTEMS Edited by Professor C.T.Leondes Published June 1990 514 pages</p> <p>The field of modern guidance and control systems has been raised to a very high level of capability because of the powerful high technology advances of the past several decades. This NATO AGARDograph captures the spirit of these powerful capabilities, and it is structured in 8 parts.</p> <p>P.T.O.</p>	<p>AGARD-AG-314</p> <p>Control equipment Guidance computers Navigational aids Flight control Air navigation Air traffic control Communication equipment Surface navigation</p>	<p>AGARDograph No.314 Advisory Group for Aerospace Research and Development, NATO ANALYSIS, DESIGN AND SYNTHESIS METHODS FOR GUIDANCE AND CONTROL SYSTEMS Edited by Professor C.T.Leondes Published June 1990 514 pages</p> <p>The field of modern guidance and control systems has been raised to a very high level of capability because of the powerful high technology advances of the past several decades. This NATO AGARDograph captures the spirit of these powerful capabilities, and it is structured in 8 parts.</p> <p>P.T.O.</p>	<p>AGARD-AG-314</p> <p>Control equipment Guidance computers Navigational aids Flight control Air navigation Air traffic control Communication equipment Surface navigation</p>
<p>AGARDograph No.314 Advisory Group for Aerospace Research and Development, NATO ANALYSIS, DESIGN AND SYNTHESIS METHODS FOR GUIDANCE AND CONTROL SYSTEMS Edited by Professor C.T.Leondes Published June 1990 514 pages</p> <p>The field of modern guidance and control systems has been raised to a very high level of capability because of the powerful high technology advances of the past several decades. This NATO AGARDograph captures the spirit of these powerful capabilities, and it is structured in 8 parts.</p> <p>P.T.O.</p>	<p>AGARD-AG-314</p> <p>Control equipment Guidance computers Navigational aids Flight control Air navigation Air traffic control Communication equipment Surface navigation</p>	<p>AGARDograph No.314 Advisory Group for Aerospace Research and Development, NATO ANALYSIS, DESIGN AND SYNTHESIS METHODS FOR GUIDANCE AND CONTROL SYSTEMS Edited by Professor C.T.Leondes Published June 1990 514 pages</p> <p>The field of modern guidance and control systems has been raised to a very high level of capability because of the powerful high technology advances of the past several decades. This NATO AGARDograph captures the spirit of these powerful capabilities, and it is structured in 8 parts.</p> <p>P.T.O.</p>	<p>AGARD-AG-314</p> <p>Control equipment Guidance computers Navigational aids Flight control Air navigation Air traffic control Communication equipment Surface navigation</p>

<p>These 8 parts are</p> <p>Part I — Integrated Guidance and Control Systems</p> <p>Part II — NAVSTAR/GPS Systems</p> <p>Part III — Optical Gyroscopes and Control Systems</p> <p>Part IV — Integrated Communication and Navigation Systems</p> <p>Part V — Integrated Navigation/Flight Control Systems</p> <p>Part VI — Civil Aircraft Navigation and Traffic Control</p> <p>Part VII — Special Topics</p> <p>Part VIII — Land Navigation Systems</p> <p>ISBN 92-835-0566-2</p>	<p>These 8 parts are</p> <p>Part I — Integrated Guidance and Control Systems</p> <p>Part II — NAVSTAR/GPS Systems</p> <p>Part III — Optical Gyroscopes and Control Systems</p> <p>Part IV — Integrated Communication and Navigation Systems</p> <p>Part V — Integrated Navigation/Flight Control Systems</p> <p>Part VI — Civil Aircraft Navigation and Traffic Control</p> <p>Part VII — Special Topics</p> <p>Part VIII — Land Navigation Systems</p> <p>ISBN 92-835-0566-2</p>
<p>These 8 parts are</p> <p>Part I — Integrated Guidance and Control Systems</p> <p>Part II — NAVSTAR/GPS Systems</p> <p>Part III — Optical Gyroscopes and Control Systems</p> <p>Part IV — Integrated Communication and Navigation Systems</p> <p>Part V — Integrated Navigation/Flight Control Systems</p> <p>Part VI — Civil Aircraft Navigation and Traffic Control</p> <p>Part VII — Special Topics</p> <p>Part VIII — Land Navigation Systems</p> <p>ISBN 92-835-0566-2</p>	<p>These 8 parts are</p> <p>Part I — Integrated Guidance and Control Systems</p> <p>Part II — NAVSTAR/GPS Systems</p> <p>Part III — Optical Gyroscopes and Control Systems</p> <p>Part IV — Integrated Communication and Navigation Systems</p> <p>Part V — Integrated Navigation/Flight Control Systems</p> <p>Part VI — Civil Aircraft Navigation and Traffic Control</p> <p>Part VII — Special Topics</p> <p>Part VIII — Land Navigation Systems</p> <p>ISBN 92-835-0566-2</p>